

RESEARCH ARTICLE

Quantifying cooperative multisite binding in the hub protein LC8 through Bayesian inference

Aidan B. Estelle¹, August George², Elisar J. Barbar^{1*}, Daniel M. Zuckerman^{2*}

1 Department of Biochemistry and Biophysics, Oregon State University, Corvallis, Oregon, United States of America, **2** Department of Biomedical Engineering, School of Medicine, Oregon Health and Science University, Portland, Oregon, United States of America

* barbare@oregonstate.edu (EJB); Zuckermd@ohsu.edu (DMZ)



Abstract

Multistep protein-protein interactions underlie most biological processes, but their characterization through methods such as isothermal titration calorimetry (ITC) is largely confined to simple models that provide little information on the intermediate, individual steps. In this study, we primarily examine the essential hub protein LC8, a small dimer that binds disordered regions of 100+ client proteins in two symmetrical grooves at the dimer interface. Mechanistic details of LC8 binding have remained elusive, hampered in part by ITC data analyses employing simple models that treat bivalent binding as a single event with a single binding affinity. We build on existing Bayesian ITC approaches to quantify thermodynamic parameters for multi-site binding interactions impacted by significant uncertainty in protein concentration. Using a two-site binding model, we identify positive cooperativity with high confidence for LC8 binding to multiple client peptides. In contrast, application of an identical model to the two-site binding between the coiled-coil NudE dimer and the intermediate chain of dynein reveals little evidence of cooperativity. We propose that cooperativity in the LC8 system drives the formation of saturated induced-dimer structures, the functional units of most LC8 complexes. In addition to these system-specific findings, our work advances general ITC analysis in two ways. First, we describe a previously unrecognized mathematical ambiguity in concentrations in standard binding models and clarify how it impacts the precision with which binding parameters are determinable in cases of high uncertainty in analyte concentrations. Second, building on observations in the LC8 system, we develop a system-agnostic heat map of practical parameter identifiability calculated from synthetic data which demonstrates that the ability to determine microscopic binding parameters is strongly dependent on both the parameters themselves and experimental conditions. The work serves as a foundation for determination of multi-step binding interactions, and we outline best practices for Bayesian analysis of ITC experiments.

OPEN ACCESS

Citation: Estelle AB, George A, Barbar EJ, Zuckerman DM (2023) Quantifying cooperative multisite binding in the hub protein LC8 through Bayesian inference. PLoS Comput Biol 19(4): e1011059. <https://doi.org/10.1371/journal.pcbi.1011059>

Editor: Ozlem Keskin, Koç University, TURKEY

Received: November 29, 2022

Accepted: March 29, 2023

Published: April 21, 2023

Copyright: © 2023 Estelle et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The code, data, and an example notebook are available at: https://github.com/ZuckermanLab/Bayesian_ITC

Funding: This work was supported by the U.S. National Institutes of Health grant R01-GM141733 (to EJB and DMZ) and by U.S. National Science Foundation Grant MCB 2119837 (to DMZ). We additionally acknowledge support in the form of computational resources of the Oregon State University NMR Facility funded in part by the National Institutes of Health, HEI Grant 1S10OD018518 (to EJB), and by the M. J.

Murdock Charitable Trust grant #2014162 (to EJB). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

Multi-site protein-protein interactions govern many protein functions throughout the cell. Precise determination of thermodynamic constants of multi-site binding is a significant challenge, however. The application of complex models to multi-step interactions is difficult and hampered further by complications arising from uncertainty in analyte concentrations. To extract binding parameter estimates from minimal and hard-to-obtain data on these large complicated systems, we utilize Bayesian statistical techniques which calculate the ‘likelihood’ of parameters giving rise to experimental observations. Bayesian computations yield probability density distributions for thermodynamic parameters of binding. To demonstrate the validity of the method and improve our understanding how the hub protein LC8 promotes dimerization of its 100+ binding partners, we test the pipeline on several of these partners and demonstrate that LC8 can bind clients cooperatively, driving interactions towards a ‘fully bound’ functional state. We additionally examine an interaction between the dimer NudE and the intermediate chain of dynein, which does not appear to bind with cooperativity. Our work provides a solid foundation for future analysis of more complicated binding interactions, including oligomeric complexes formed between LC8 and clients with multiple LC8-binding sites.

Introduction

Intracellular processes frequently depend on complex, multistep interactions between proteins or between proteins and small-molecule ligands [1–3]. The hub protein LC8 provides an extreme example of binding complexity, accommodating over 100 client proteins via two symmetrical binding grooves [4,5]—often binding in multivalent fashion with a range of stoichiometries [6–10]. LC8 is found throughout the eukaryotic cell and is involved in a host of cell functions, with client proteins including transcription factors [7,9], tumor suppressors and oncogenes [11,12], viral proteins [13–15], and cytoskeletal proteins [6,16].

Structurally, LC8 forms a small 20 kDa homodimer (Fig 1A), with two identical binding grooves formed at the dimer interface [4,5]. These binding sites induce a beta-strand structure in a well-characterized linear motif anchored by a TQT amino acid sequence within disordered regions of client proteins [6,9]. Despite extensive study [9,16,17], the mechanistic thermodynamics of LC8 binding are still not fully understood, due to the difficulty of deconvoluting a multiplicity of microscopic states in its binding processes.

While published isothermal titration calorimetry (ITC) experiments fit LC8-client interactions to a simple model [9,16,17], binding in fact occurs in two distinct steps (Fig 1A). Evidence of this was first observed in nuclear magnetic resonance (NMR) titrations of peptides into LC8, where, for some clients, a partially bound intermediate state was observable. Populations for each state were fit to a two-site model, which suggested weak cooperativity in binding, although statistically rigorous investigation was not undertaken [18]. Further evidence that LC8 binding cannot be explained by a simple model emerged from ITC experiments that exhibit non-sigmoidal behavior, where rather than forming a plateau, early injections dip in heat per injection, forming a U shape at the beginning of the titration (Fig 1B) [9]. Single-site models, such as the identical sites model in Origin software [19], return strictly sigmoidal binding curves and do not fit the U shape in such isotherms. Although isotherms of LC8 binding partners [9] were fit to an identical sites model, that study did not probe details of binding thermodynamics. The observed non-sigmoidal behavior clearly raises the possibility that these

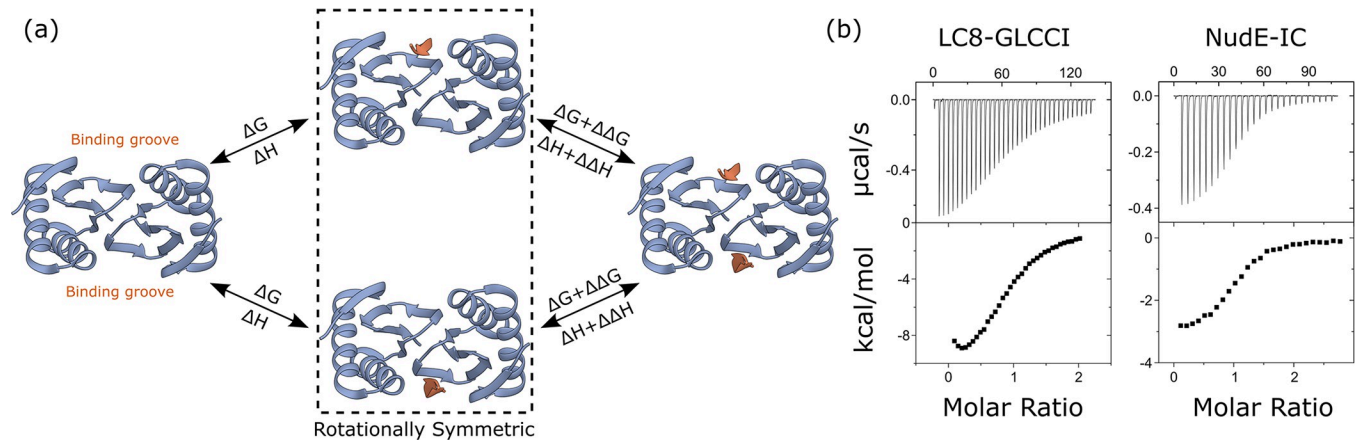


Fig 1. LC8 binds clients through a two-site mechanism. (a) Diagram of LC8-client binding, showing a structure of apo LC8 on the left, and a fully bound structure (PDB 3E2B) on the right. Intermediates are boxed to indicate they are symmetric and indistinguishable species. (b) Example isotherms for binding between dimeric LC8 and client peptide taken from GLCCI (left) and between the coiled-coil protein NudE and its client, the intermediate chain (IC) of dynein (right).

<https://doi.org/10.1371/journal.pcbi.1011059.g001>

isotherms may fit well to a two-site model of binding more representative of expectations for LC8's two client sites [20].

The use of ITC to interrogate complex systems and multisite binding is challenging, as ITC data are of relatively low information, and individual isotherms often fit well to varied model parameters [20,21]. Despite this, ITC experiments can measure cooperativity [22,23], entropy-enthalpy compensation [17,24], changes in protonation state [25,26], and competition between multiple ligands [27,28]. In general, these studies rely on fitting data globally to a model that includes several isotherms collected at varied conditions to reduce ambiguity of fit parameters [20,21], or a 'divide and conquer' type approach, where subsections of a complex binding network can be isolated and examined [16,22].

Concentration uncertainty is a critical concern in analysis of ITC data. In principle, accurate determination of protein and ligand concentration is a prerequisite for obtaining reliable thermodynamic quantities, yet these values are challenging to obtain for many systems [29–33]. The most common software package for fitting ITC data, built into the data analysis and fitting software Origin, and distributed with calorimeters, attempts to account for this uncertainty in several models through the stoichiometric parameter n , which can fit to non-integer values to correct for error in macromolecule concentrations [19,34]. However, in addition to assuming independent sites, this implementation ignores uncertainty in concentration of the titrant in the syringe, which is treated as a fixed value. The popular and highly flexible fitting software SEDPHAT greatly improves on Origin's capabilities, allowing for both explicit or implicit (i.e. an 'inactive fraction' correction) uncertainty corrections [21,35]. As the authors note, however, allowing for variation in both analyte concentrations makes binding constants indeterminable within SEDPHAT due to correlative effects among model parameters.

Bayesian analysis offers a natural framework for incorporating uncertainty in concentration measurements in ITC analysis [32,36,37], and has been successfully applied to measurements of binding using both ITC and other methods [36,38,39]. In a Bayesian framework, thermodynamic parameter determination is guided by a mix of experimental data and 'prior' information, such as uncertainty ranges/models, that weights the overall 'posterior' probability of a given set of thermodynamic parameters. The posterior distribution of estimated binding parameters generated through Bayesian analysis is a complete description of the probability range of each model parameter—and correlations among parameters—based on the input data

and priors. With a meaningful prior description of concentration uncertainty, there is reduced risk of underestimating uncertainty in thermodynamic binding parameters. The Bayesian framework accounts for the full-dimensional likelihood of parameter space by construction, in contrast to maximum-likelihood approaches to uncertainty quantification in multi-parameter systems which approximate the likelihood function based on optimal parameters [21,40].

We build on earlier applications of Bayesian inference to ITC. Nguyen et al. [32] studied 1:1 binding using a Bayesian statistical framework accounting for concentration uncertainty and performed sensitivity analysis on concentration priors. Very recently, Nguyen et al. [37] further explored systems of high uncertainty, with a focus on determining multiple binding affinities from ITC experiments on enantiomeric mixtures. The work emphasized the importance of concentrations in uncertainty determination, finding that different priors for concentrations impacted the accuracy of confidence intervals and demonstrated that in some cases multiple distinct affinities can be determined accurately from a single isotherm. For a two-site binding model, Duvvuri et al. [39] demonstrated that a Bayesian method can accurately and precisely determine affinities in two-site models when using global modelling of several isotherms, but the work assumes no uncertainty in measured concentrations [39], raising the possibility that parameter uncertainty is underestimated [21,32]. Cardoso et al. [36] used a simplified 4-site binding model with a single common binding enthalpy for a set of isotherms to determine 3 of 4 distinct affinities between protein and ligand, with the fourth being uncertain across a range of several orders of magnitude. Although Cardoso et al. [36] include concentrations as model parameters, they greatly narrow concentration priors using a preliminary ‘calibration’ assuming identical sites. We note that such a model is not appropriate for complex systems, particularly in cases where the identical-sites model does not fit well to the isotherm shape. A sensitivity analysis regarding concentration uncertainty was not performed in either multisite study, and neither work probed the information content of single isotherms for multisite systems.

Here, we report a Bayesian analysis of two-site systems with proper accounting of concentration effects critical for reliable analysis. We show that LC8-client interactions unambiguously exhibit positive cooperativity, driving binding towards a fully bound state. In contrast, symmetric two-site binding between the coiled coil domain of the dynein cargo adaptor NudE and the intermediate chain (IC) of dynein [41] shows no significant evidence for cooperativity.

We also provide methodological advances. First, we derive simple mathematical relations that govern the influence of concentration uncertainties on different binding parameters, providing a fundamental basis for the previously noted strong sensitivity of enthalpies—but not free energies—to concentration uncertainty [21]. Second, by using synthetic models, we systematically characterize the causes of binding-parameter uncertainties in two ways: we demonstrate that substantial uncertainty can result from mismatch between binding parameters and experimental conditions; and we also determine the effects of different prior functional forms and uncertainty ranges in a multisite context, extending the work of Nguyen et al. [32]. Finally, we outline best practices for determining model parameters and uncertainties in a multisite Bayesian framework.

Results

A mathematical “degeneracy” in thermodynamic parameters impacts analysis at any stoichiometry

We first present a simple mathematical analysis that explains previously reported correlation effects among titrant and titrand concentrations [21], and which significantly impacts the overall analysis of ITC data. Specifically, when the concentrations are uncertain, as is common

in analysis of ITC data [21,32], we show below that only the *ratio* of titrant:titrand concentrations can be estimated, rather than the individual values, and this ambiguity propagates to all thermodynamic parameters. Hence, there is a “degeneracy” in that multiple solutions (sets of concentration values and thermodynamic parameters) will equally describe even idealized ITC data lacking experimental noise (Fig 2).

We first describe the degeneracy for standard 1:1 binding between a macromolecule M and ligand X, following the scheme



The heat, Q, of a 1:1 binding system at any titration point can be described using the standard quadratic binding equation used in the identical sites model [19,42]:

$$\frac{Q}{V_0} = \frac{[M_t]\Delta H}{2} \left\{ 1 + \frac{[X_t]}{[M_t]} + \frac{K_d}{[M_t]} - \sqrt{\left(1 + \frac{[X_t]}{[M_t]} + \frac{K_d}{[M_t]} \right)^2 - \frac{4[X_t]}{[M_t]}} \right\} \quad (2)$$

where $[M_t]$ and $[X_t]$ are the concentrations of macromolecule and ligand (i.e., cell component and syringe component) respectively, while K_d and ΔH are the binding affinity and enthalpy. The degeneracy is demonstrated by introducing a linear scaling of all parameters by an arbitrary number denoted α . Specifically, we apply the following transformations:

$$[M_t] \rightarrow \alpha[M_t]$$

$$[X_t] \rightarrow \alpha[X_t]$$

$$K_d \rightarrow \alpha K_d$$

$$\Delta H \rightarrow \frac{\Delta H}{\alpha} \quad (3)$$

Applying this set of transformations, we can rewrite the binding equation:

$$\frac{Q}{V_0} = \frac{\alpha[M_t]\frac{\Delta H}{\alpha}}{2} \left\{ 1 + \frac{\alpha[X_t]}{\alpha[M_t]} + \frac{\alpha K_d}{\alpha[M_t]} - \sqrt{\left(1 + \frac{\alpha[X_t]}{\alpha[M_t]} + \frac{\alpha K_d}{\alpha[M_t]} \right)^2 - \frac{4\alpha[X_t]}{\alpha[M_t]}} \right\} \quad (4)$$

Regardless of the value of the factor α , all introduced factors cancel leaving Q unchanged.

We emphasize that the factor α is used here as a mathematical tool to analyze the governing equations, but α is *not* a parameter in our Bayesian inference pipeline, and we do *not* assume analyte concentrations bear any fixed relation to one another in the inference pipeline. In the Bayesian inference process, parameters can take any values within the ranges allowed by the priors.

Nearly identical formal considerations apply in the two-site binding model of primary interest here. As detailed in the Methods, the value of Q is unchanged when both concentrations and both K_d values are multiplied by α and both ΔH values are divided by α . The underlying model is more complex as it requires solving a system of nonlinear equations (see [Methods](#) for details), but the result is that α is propagated through the nonlinear equation solutions, and once again cancels in the calculation of Q, leaving the heat value unchanged.

To facilitate analysis and discussion of cooperativity below, we parameterize our two-site model using ΔG , $\Delta\Delta G$, ΔH and $\Delta\Delta H$. Values are ‘microscopic’ terms corresponding to a single site-resolved binding step. The $\Delta\Delta G$ and $\Delta\Delta H$ value correspond to the differences between the

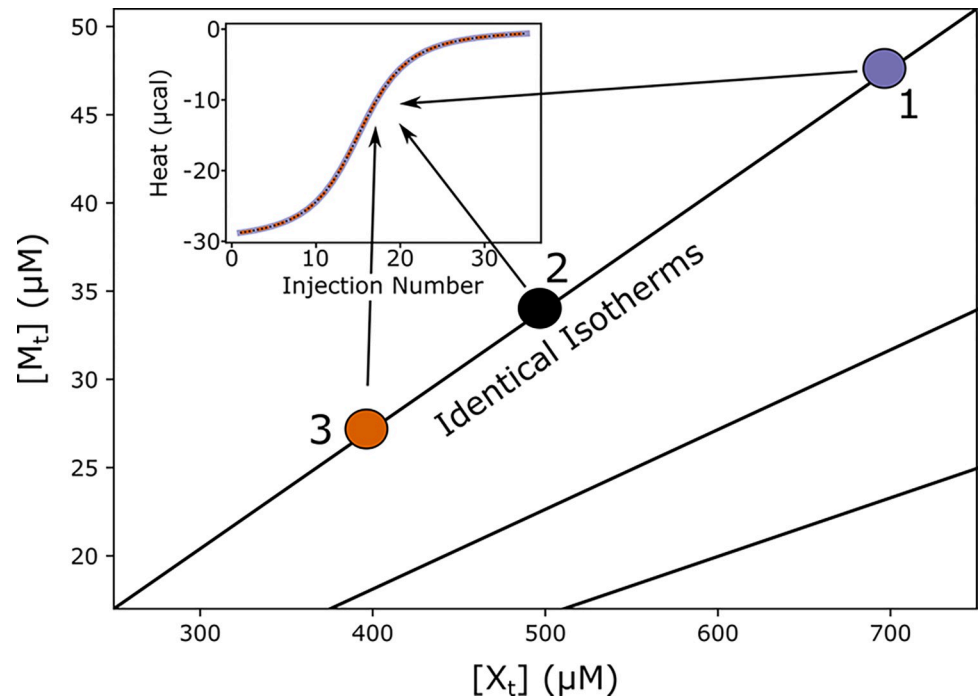


Fig 2. Exact degeneracy in binding isotherms. Based on the scaling relations of Eq (3), for any set of ligand and total macromolecule concentrations (X_t , M_t), there are infinitely many alternative concentrations (e.g., filled circles) on a diagonal line in the ($[X_t]$, $[M_t]$) plane which yield exactly equivalent isotherms (inset, isotherms for points 1, 2, and 3 are drawn in distinct colors but overlay exactly) for a fixed set of thermodynamic parameters. For any given point in parameter space, equivalent degenerate lines can be drawn in a radial manner (e.g. the two additional black lines), passing through the point and the origin. The plotted synthetic isotherms are for 1:1 binding, but analogous degeneracy also holds for multivalent binding—see text. Note that no fixed relationship among concentrations is assumed during Bayesian inference.

<https://doi.org/10.1371/journal.pcbi.1011059.g002>

first and second microscopic binding steps. Thus $K_{d1} = e^{\Delta G/RT}$, $K_{d2} = e^{(\Delta G + \Delta \Delta G)/RT}$, and $\Delta H_2 - \Delta H_1 = \Delta \Delta H$. The energy-like formulation allows for easy assessment of cooperativity ($\Delta \Delta G$ will be zero in the absence of cooperativity and positive or negative for negative or positive cooperativity, respectively), and $\Delta \Delta H$ is the change in enthalpy between binding steps with analogous characterization.

The degeneracy and associated scaling relationships in Eq (3) provide important insight into assessment of thermodynamic parameters inferred from ITC data. We see directly that binding enthalpy changes proportionately to concentrations of titrant and titrand. That is, a given percent error in an assumed concentration of either ligand (characterized by α) translates to the same scale of error in ΔH . On the other hand, the binding free energy ΔG , is less sensitive to concentration errors, due to scaling with $\ln(\alpha)$, rather than being directly multiplied by α . We note again that α is used only in the formal analysis here and not in our Bayesian inference process where parameters are not assumed to have any fixed relation with one another.

The scaling relationships of Eq (3) also presage a significant issue in Bayesian inference, namely, sensitivity to the choice of priors. Within the set of degenerate solutions (diagonal lines of concentration pairs in Fig 2), the Bayesian ‘likelihood’ probability—which describes how well a parameter set fits the data in the absence of prior information—will be constant, as solutions are mathematically identical. Thus, within any degenerate set, the assumed prior distributions for concentrations will determine the overall posterior distributions (see Methods).

Because the posterior distributions ultimately determine the uncertainty ranges, this is a key point.

Below, we continue to examine the ramifications of the concentration degeneracy, demonstrating concretely that enthalpy is more impacted by uncertainty in concentrations than free energy. We also examine the influence of priors on parameter distributions and discuss parameter distributions determined from isotherms in cases of high concentration uncertainty.

Validation of Bayesian inference pipeline with synthetic data

To test our Bayesian pipeline ([Methods](#)), we generated ‘synthetic’ simulated isotherms using hand-chosen sets of thermodynamic parameters ΔG , $\Delta\Delta G$, ΔH , $\Delta\Delta H$ (see [Fig 1](#)) inserted in [Eq 17](#) with added Gaussian noise. Following an exploration using synthetic data of how cooperativity impacts binding isotherms (e.g. [Fig 3A](#)), we selected synthetic model parameters to mimic the isotherm shape seen in LC8-peptide binding examples. Specifically, slight positive cooperativity ($\Delta\Delta G = -1$, $\Delta\Delta H = -1.5$ kcal/mol) was best-suited to imitating real LC8-peptide isotherms, along with $\Delta G = -7$ and $\Delta H = -10$ kcal/mol. Synthetic noise is taken from a Gaussian distribution with a zero mean and standard deviation $\sigma = 0.2$ μ cal. As shown in [Fig 3](#), we used our pipeline to sample posterior distributions for these isotherms. For concentrations, we chose uniform prior distributions within $\pm 10\%$ of the true value (which simply limits sampled concentration values to these ranges). The choice of 10% approximates what we view to be an attainable level of uncertainty for experimental protein concentrations.

Under these representative conditions, inferred posterior distributions fell around the known model parameters, and model parameters equate to isotherms which closely matched the isotherm shape ([Fig 3B and 3C](#)). The finite widths of the distributions are due both to synthetic experimental noise and correlative effects from the discussed model degeneracy. The posterior distribution for ΔG covers a range of ~ 1 kcal/mol distributed around the true value of -7 kcal/mol. Examination of the distribution lets us define a ‘credibility region,’ that contains 95% of the distribution probability (i.e., from the 2.5 to 97.5%ile of the distribution), which is directly analogous to a confidence interval in frequentist terms. For ΔG , the 95% credibility region is -7.5 to -6.4 kcal/mol. Similarly, the 95% credibility region for $\Delta\Delta G$ covers a range of ~ 1.5 kcal/mol, evenly distributed around -1 kcal/mol. ΔH and $\Delta\Delta H$ both have slightly wider credibility regions, with widths of 2.3 and 3.3 kcal/mol respectively, but both are distributed around the true values of -10 and -1.5 kcal/mol respectively.

In a “bottom line” assessment of the Bayesian credibility regions, we used synthetic data to check the percent of times the true value was covered based on multiple independent trials. In particular, we modeled 100 replicate isotherms each with different measured concentrations chosen from priors and different synthetic noise. For each parameter, we compared the frequency with which a given interval included the true value against the expectations for the interval [[32,37,38](#)]. For example, a 50% credibility interval should include (“cover”) the ground-truth value in 50% of the replicates. Plotting the expected credibility against the actual coverage for our replicates ([S2 Fig](#)) returns a nearly straight line, indicating the model is producing accurate credibility regions and does not significantly under- or over-estimate parameter uncertainty.

One benefit of Bayesian inference is the ability to examine multi-dimensional likelihood distributions to obtain correlations between model parameters without approximation. For example, in our two-dimensional distributions for the thermodynamic parameters, the ΔG and $\Delta\Delta G$ values are strongly negatively correlated ([Fig 3C](#)), indicating a compensatory effect in the model, where increases in ΔG can be compensated by decreases in $\Delta\Delta G$ to arrive at

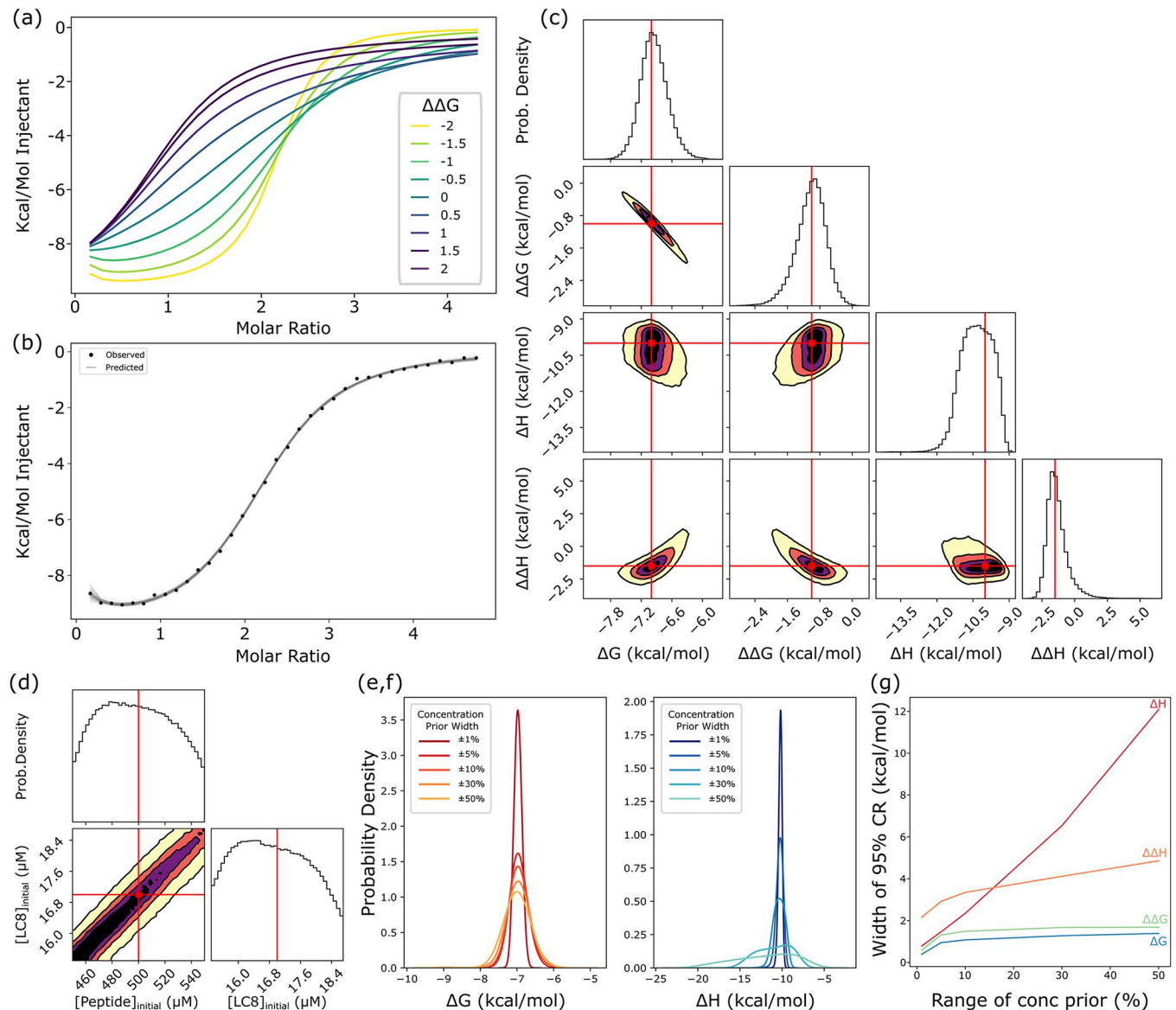


Fig 3. Analysis of two-site model using synthetic isotherms. (a) A set of synthetic isotherms for two-site binding with varied $\Delta\Delta G$ parameters demonstrating how cooperativity changes isotherm shape. Thermodynamic parameters are $\Delta G = -7$, $\Delta H = -10$, and $\Delta\Delta H = 0$. Concentrations are set at 17 and 500 μM for cell and syringe respectively, and injection volumes are 6 μL . (b) A synthetic isotherm ($\Delta G = -7$, $\Delta H = -10$, $\Delta\Delta G = -1$, $\Delta\Delta H = -1.5$ kcal/mol) with added Gaussian noise (points) with 50 fitted isotherms (lines) generated through the Bayesian pipeline, i.e., sampled from the posterior. (c) One and two-dimensional marginal distributions for thermodynamic parameters, with contours in the two-dimensional plots set at 95 (yellow), 75 (orange), 50 (purple) and 25% (black) confidence. Red lines and dots indicate true values for the synthetic isotherm. Marginal distributions and MCMC chains for all eight model parameters, including nuisance parameters can be found in S1 Fig. (d) Marginal distributions for concentration parameters, exhibiting characteristic diagonal shape (Fig 2) with contours as in (c). Plots at the top of each column in panels c and d are one-dimensional probability density distributions. (e,f) One-dimensional distributions for ΔG (e) and ΔH (f) plotted for models with prior ranges for concentrations of 1, 5, 10, 30 and 50% of the stated concentration. (g) Width of the 95% Bayesian credibility region, akin to a confidence interval, for thermodynamic parameters as a function of the width of the concentration prior used in modeling, plotted from models with prior ranges for concentrations of $\pm 1, 5, 10, 30$ and 50% of the stated concentration.

<https://doi.org/10.1371/journal.pcbi.1011059.g003>

similar solutions. Resultantly, the distribution for both ΔG and $\Delta\Delta G$ are broader than the 'total' free energy (i.e., $2\Delta G + \Delta\Delta G$), evidence the overall energy of binding can be determined more precisely than the energy of each step (S3 Fig). Additionally, the mathematical degeneracy for concentrations described above can clearly be seen in these two-dimensional

correlations: the two-dimensional marginal distribution for each concentration is a straight line of a width determined by noise, covering the entire prior range (Fig 3D). Similar correlations can also be seen in two-dimensional distributions for the concentrations plotted against thermodynamic parameters (S1 Fig), especially in the ΔH term. The scaling relationship of the model parameters outlined previously means that each point along these diagonals corresponds to a degenerate solution, i.e., each point has equivalent likelihood based on the data.

Impact of concentration degeneracy on two-site thermodynamic parameters assessed via synthetic data

Bayesian inference enables determination of distributions for thermodynamic parameters even in cases of a concentration degeneracy. The net result, as will be seen, is a broadening of (posterior) parameter distributions based on multiple equally likely solutions, constrained by the priors used. Despite intrinsic limitations surrounding concentrations, the *ratio* of concentrations can be quantified with relatively high precision even when individual concentrations are highly uncertain.

To quantify the impacts of the concentration degeneracy within a Bayesian inference pipeline, we examined a series of uniform prior distributions for concentrations, ranging from $\pm 1\%$ to $\pm 50\%$ for both concentrations. These priors were applied to a synthetic isotherm mimicking experimental parameters, as described in the pipeline validation above. The choice of concentration priors—which embody assumed or estimated experimental uncertainties—greatly impacts the predicted uncertainty of thermodynamic parameters. The distributions for ΔG and ΔH , not surprisingly, both widen as the prior range is increased (Fig 3E and 3F). As anticipated by the degeneracy scaling relations of Eq (3), the width of the distributions for ΔH and $\Delta\Delta H$ increases roughly linearly with the concentration prior range, while the distributions for ΔG and $\Delta\Delta G$ initially increase linearly at low concentration ranges then increase more slowly (Fig 3G). This can be explained by the logarithmic relationship between the K_d (which is what scales with the degeneracy) and free energy. Functionally, high uncertainty in concentrations therefore only slightly increases uncertainty in binding free energy, while having a more significant impact on binding enthalpy. This is also apparent when examining two-dimensional distributions for concentrations plotted against thermodynamic parameters, where ΔH can be seen to correlate clearly with concentrations (S1 Fig).

The concentration degeneracy of the model limits the degree to which erroneously determined individual concentrations can be corrected. As discussed above, the fact that the Bayesian likelihood is uniform at any point along the degeneracy lines (Fig 2) means that the data have little impact on the posterior distributions for *individual* concentrations, which instead takes the shape of the prior used. This can be seen in the model validation example (Fig 3D), where the posterior distribution is approximately uniform, echoing the uniform prior.

The ratio of concentrations ('macromolecule' to 'ligand'), on the other hand, is a determinable parameter, as the ratio does not change along the degenerate line. Posterior distributions are therefore limited to this ratio. For example, when we sample the posterior for the same isotherm, but use a normal (i.e. Gaussian) distribution for one concentration prior and a uniform distribution for the other, both posteriors take the shape of a normal distribution (S4 Fig). This is a direct result of the degeneracy identified above. S1 Table shows concentration ratio credibility regions for the experimental systems. Because of the nearly determinative relationship between the prior and posterior concentration distributions, we elected to use uniform priors for concentrations throughout this work to avoid undue influence on our results from model priors.

For completeness, we also examined 1:1 binding with synthetic data. Overall, the impact of the concentration degeneracy on model parameters is similar (S5 Fig): binding enthalpy posterior distributions are wider than free energy distributions. In response to changes in concentration prior ranges, the posterior for ΔG is more impacted than in the two-site model, but the distribution remains much narrower than that of the enthalpy, as in the two-site model.

Application to 2:2 LC8:IDP Systems

We applied the Bayesian analysis pipeline to a set of 7 experimental isotherms of binding between LC8 and client peptides, all of which bind in a 2:2 ratio. Note that the two LC8's form a strong homodimer ($K_d \sim 60$ nM)[43] and this initial homodimer formation is excluded from our analysis. Client peptides were chosen based on a K_d of less than 5 μ M when fit to an identical-sites model, and deviation from the standard sigmoidal isotherm shape [9]. As noted above, the user-supplied uncertainties for concentrations may impact uncertainty in other parameters. Following analysis with priors of $\pm 10\%$ and $\pm 20\%$ of the measured LC8 concentration as determined by absorbance at 280 nm, we have elected to focus on results at $\pm 10\%$ (Table 1), as using $\pm 20\%$ does not greatly alter the posterior distributions (S2 Table). The high degree of purity of LC8 (>95%) and high absorbance at 280 nm, due to the presence of 6 chromophores (1 Trp, 5 Tyr) allow for a high signal-to-noise ratio for the absorbance, reducing uncertainty in the measurement. Comparatively, because of the difficulty in accurately measuring concentration for peptides with few or no chromophores [30,44] (1 Tyr residue for the peptides discussed here [9]), we used a prior of increased width for the peptide concentration, up to a limit of $\pm 50\%$ of the initially measured value estimated by absorbance at 280 nm. As discussed above, the posterior distributions processed through the Bayesian pipeline are limited by the most restrictive prior used, owing to the concentration ratio being well defined (S2 Table). As a result, this approach ensures that posterior distributions are limited to the range around the measured concentration of LC8, allowing us to effectively infer the uncertain peptide concentration.

Bayesian analysis of the seven systems reveals significant heterogeneity in the precision with which binding parameters can be determined (Table 1). As will be described in detail below, this is only partially reflective of apparent data quality (e.g., noise level). Instead, certain binding parameters, particularly binding enthalpies, are intrinsically more difficult to characterize. Variations in precision do not stem from inadequate sampling in the Bayesian pipeline: triplicate runs are performed to confirm sampling quality (see Methods) (example in S6 Fig).

In particularly tractable cases, such as for SPAG5 binding in Fig 4, the analysis provides marginal distributions of similar precision to those seen with synthetic data. For binding between a peptide from the protein SPAG5 and LC8, Bayesian analysis yields a 95% credibility

Table 1. Ranges for thermodynamic parameters for LC8-client binding. Values delineate 95% Bayesian credibility regions from sampled posterior distributions in kcal/mol, which are akin to 95% confidence intervals. Previously published binding parameters from an identical-sites model for these isotherms are available in S3 Table.

Peptide	ΔG		$\Delta\Delta G$		ΔH		$\Delta\Delta H$		-TAS		-T Δ AS	
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
SPAG5	-6.9	-6.2	-2.1	-1.1	-18	-14	-1.8	3.6	6.9	11	-5.7	0.6
BSN (I)	-5.6	-4.9	-2.1	-0.9	-37	-14	0.1	45	8.8	32	-51	-2.6
BSN (II)	-7.1	-6.5	-1.2	-0.4	-5.8	-4.8	-9.1	-7.0	-1.9	-1.0	6.1	8.3
SLC9A2	-6.8	-5.5	-2.6	-0.5	-24	-10	-5.0	23	3.7	18.3	-26	4.4
VP35	-7.4	-6.8	-1.7	-0.9	-14	-11	-0.7	1.5	4.0	6.7	-3.0	-0.2
GLCCI	-6.1	-5.1	-2.6	-1.0	-27	-9.7	-0.5	36	3.7	22	-39	-0.5
BIM	-9.5	-7.1	-2.0	0.8	-12	-10	-0.5	2.9	1.4	4.4	-0.7	2.2

<https://doi.org/10.1371/journal.pcbi.1011059.t001>

region of -6.9 to -6.2 kcal/mol for ΔG (Table 1), equivalent to a range for K_{d1} of 8.7 μM to 27 μM . The 95% credibility region for $\Delta\Delta G$, the allosteric difference between the first and second binding event, is -2.1 to -1.1 kcal/mol, roughly equivalent to a 6 to 30-fold increase in affinity for the second binding step relative to the first. The change in binding enthalpy between first and second events, $\Delta\Delta H$, is distributed around zero (Fig 4B), with uncertainty >2 kcal/mol for all cases, meaning we are unable to discern conclusively if there is any change in enthalpy between binding steps. From ΔG and ΔH values for both binding steps, we can additionally calculate $-T\Delta S$ and $-T\Delta\Delta S$, for the entropy of binding and the change in entropy across binding steps respectively. Although the marginal distributions for these terms are broad (Fig 4D), the $-T\Delta\Delta S$ mostly sits at negative values, indicating that binding enhancement has a greater probability of being entropically driven. See Table 1 for the full set of credibility regions.

Some general conclusions about cooperativity are apparent from the full set of data (Table 1). In all cases except one (binding to BIM), the distribution for $\Delta\Delta G$ is negative, indicating that all isotherms exhibit some positive cooperativity. Even for BIM, which has the widest $\Delta\Delta G$ distribution, the range predominantly covers negative values. All isotherms exhibit precisely determined free energies: 95% credibility regions cover a range of 2 kcal/mol or less for all cases except BIM. A common feature among some isotherms, seen clearly in the middle and right examples in Fig 5, is an apparent loss of precision in our ability to determine model enthalpies, as both show wide distributions for ΔH and $\Delta\Delta H$. For these isotherms (e.g., SLC9A2, GLCCI, and BIM), the two-dimensional marginal distribution for ΔH and $\Delta\Delta H$ shows a clear correlative effect (S7 Fig.), and the one-dimensional distribution for the 'total' enthalpy (i.e. $2\Delta H + \Delta\Delta H$) is narrower than the individual parameter distributions (S1 Table). In sum, the wide enthalpy distributions represent an inability to precisely determine

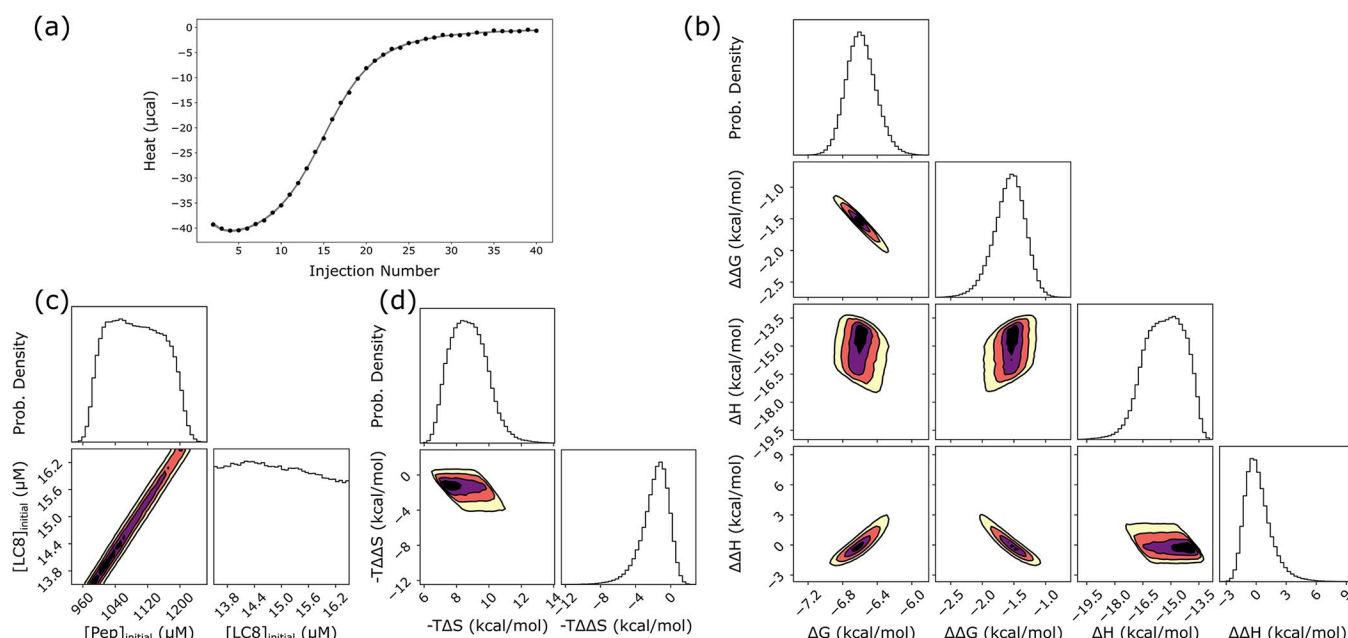


Fig 4. LC8 binding to a peptide from the protein SPAG5. (a) Experimental titration isotherm of SPAG5 into LC8 (points) with 50 example traces (lines) drawn from the posterior distribution of thermodynamic parameters and concentrations. (b) One and two-dimensional marginal distributions for thermodynamic parameters, with contours in the two-dimensional plots set at 95 (yellow), 75 (orange), 50 (purple) and 25% (black) credibility. (c) Marginal distributions for concentrations of LC8 and peptide, showing a line of degenerate solutions, which may be compared to Fig 2. (d) Marginal distributions for entropy ($-T\Delta S$) and change of entropy ($-T\Delta\Delta S$). Plots at the top of each column in panels b,c,d are one-dimensional probability density distributions.

<https://doi.org/10.1371/journal.pcbi.1011059.g004>

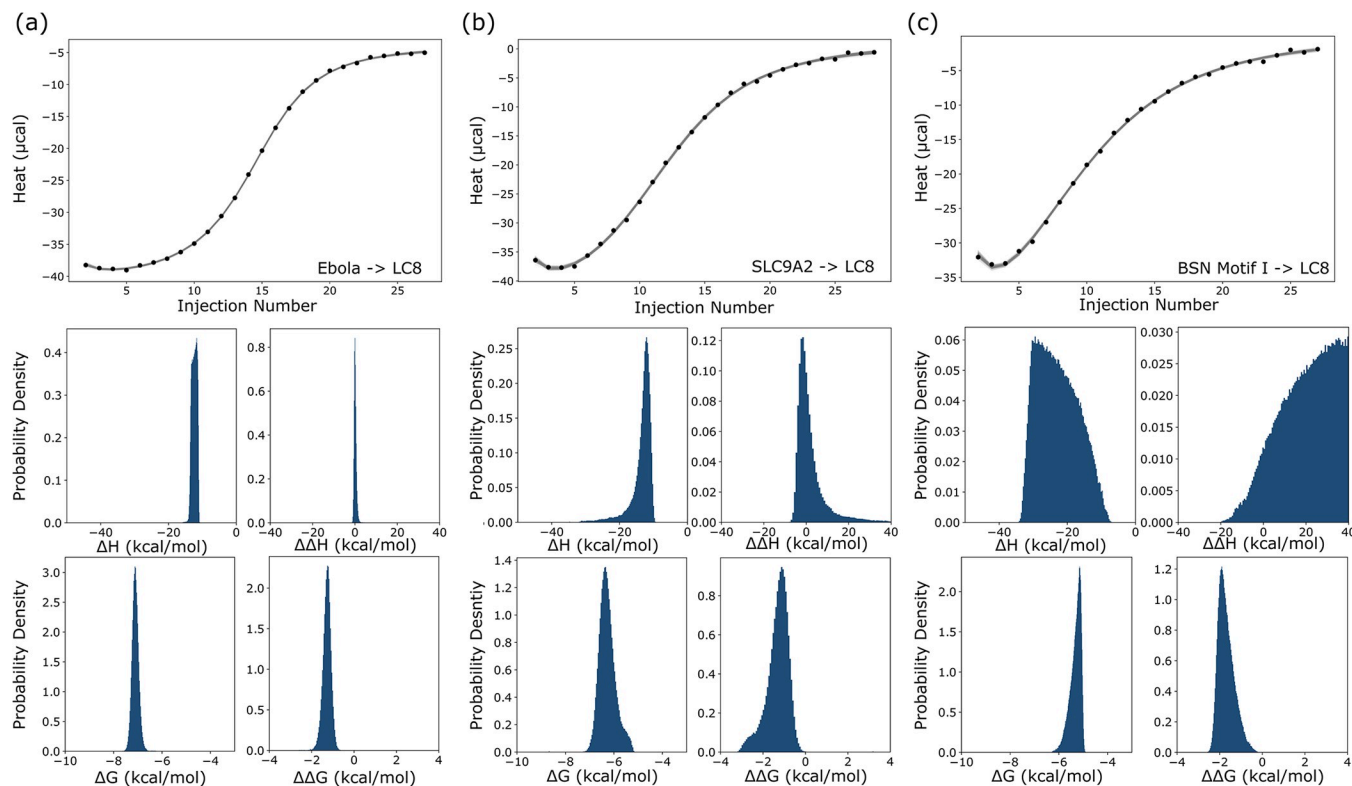


Fig 5. Thermodynamic parameter distributions from 3 LC8-peptide isotherms. Binding between LC8 and peptides from Ebola VP35 (a), SLC9A2 (b) and motif 1 from BSN (c). Isotherms are shown at the top, and distributions for thermodynamic parameters are shown below. Horizontal axes represent the full width of the uniform prior range for each parameter to allow for direct comparison between each isotherm.

<https://doi.org/10.1371/journal.pcbi.1011059.g005>

‘microscopic’ enthalpies for individual binding events. Nevertheless, even in these cases, the overall enthalpy can be determined with high precision (S1 Table).

Parameter inference from multiple isotherms

The use of additional experimental information is expected to increase the precision of parameter determination, and Bayesian inference is readily adapted to employ multiple isotherms, whether at matching or different experimental conditions [39]. Despite the higher dimensionality resulting from additional nuisance parameters (see Methods), we found it relatively easy to sample the parameter space for a two-site model including two-isotherms for several of our LC8-client interactions (S8 Fig). For GLCCI, for example, the addition of a second isotherm narrowed posterior distributions, while in others (e.g. BSN motif I) it proved less impactful, largely just taking the same shape as the distribution for individual isotherms. We note that the isotherms examined were designed as technical replicates, not as optimized isotherms at different conditions for a global model. We expect results on multiple isotherms with varied experimental setups, e.g., different concentrations, to be more consistently valuable. Nevertheless, the global models demonstrate our ability to apply the pipeline to multiple isotherms simultaneously, a key step toward improved precision going forward.

NudE-IC binding

To confirm the utility of the Bayesian pipeline for a range of proteins with two sites, we tested it on binding between the coiled-coil dimer NudE and the intermediate chain (IC) of dynein.

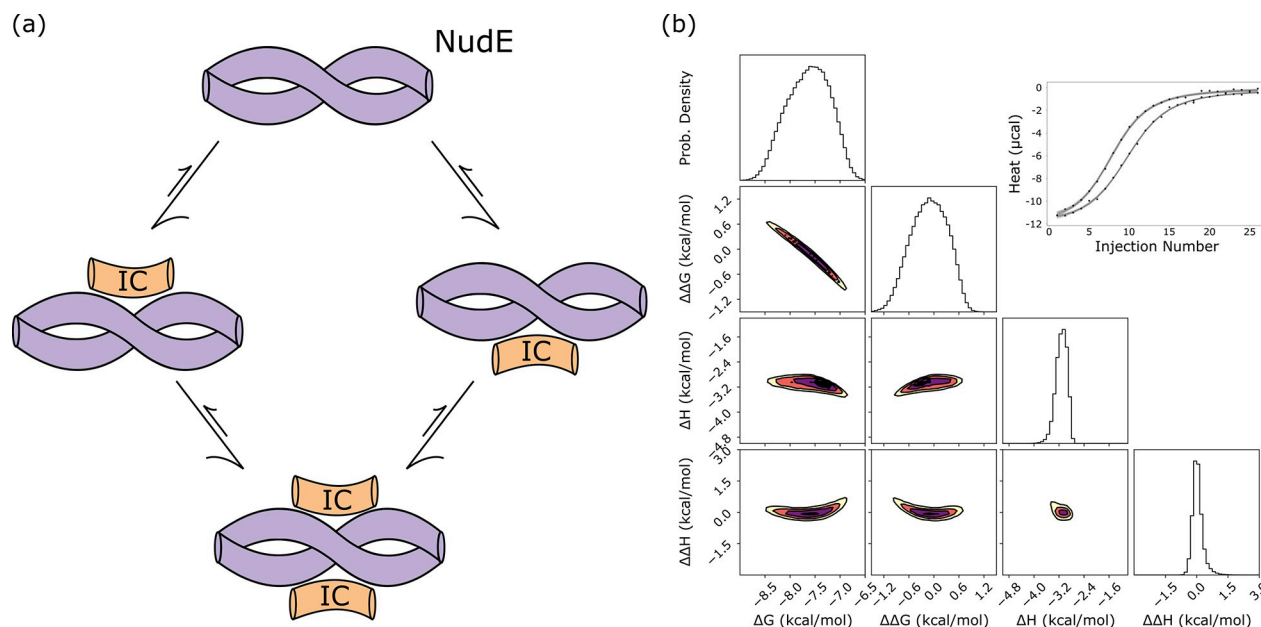


Fig 6. binding between the intermediate chain (IC) and NudE. (a) A model of NudE-IC binding, which forms a 2:2 complex. A cartoon diagram of NudE is shown in purple and IC in orange. (b) Sampled distributions modeled from a 'global' model based on two isotherms for binding between IC and NudE from yeast. Marginal distributions for thermodynamic parameters are shown on the left, and the top right corner contains the experimental isotherms (points) with model values (lines) drawn from the posterior.

<https://doi.org/10.1371/journal.pcbi.1011059.g006>

Binding between NudE and IC can be described by the same model as binding between LC8 and clients—NudE forms a strong ($K_d \sim 200$ nM in *C. thermophilum* [45]) dimeric coiled-coil structure which then accommodates two chains of monomeric disordered IC for a 2:2 complex stoichiometry (Fig 6A)[41]. Prior characterization of NudE-IC binding used a simple identical-sites binding model with a single K_d , and thus provides a good system for re-analysis as well as for comparison to LC8-client binding [41,46].

For NudE-IC binding, a two-site model recapitulates the parameters determined in fits to identical sites modeling using Origin. For high confidence in model parameters, we applied a global model, identical to the one used in LC8-client binding, to two titrations of IC into NudE. Bayesian sampling returns narrow distributions for all thermodynamic parameters, both for individual-isotherm models (S9 Fig), and for the global, 2-isotherm model (Fig 6). Neither $\Delta\Delta G$ nor $\Delta\Delta H$ are significantly shifted from a distribution around zero, suggesting little, if any, cooperativity in binding. Published work applying an identical-sites model to these data provides a binding enthalpy of -3.1 kcal/mol, and an affinity of 2.3 μ M (i.e. a ΔG of -7.6 kcal/mol) implying a $T\Delta S$ value of 4.5 kcal/mol [47]. Our two-site model predicts ΔH and ΔG distributions centered near -3 kcal/mol, and -7.5 kcal/mol respectively, aligning well with the published values. This binding interaction works well as a counterexample to LC8-client binding: distributions for allosteric terms are centered around zero and determined distributions match closely to reported values modeled from a simple model.

Limits of precision in binding enthalpies

We exploit synthetic isotherms to systematically survey binding parameters and determine the extent to which the physical parameters intrinsically lead to lower precision in parameter inference. This effort was motivated by the disparity between the three example isotherms in Fig 5 and anecdotal observations that weaker binding, such as between LC8 and BSN I

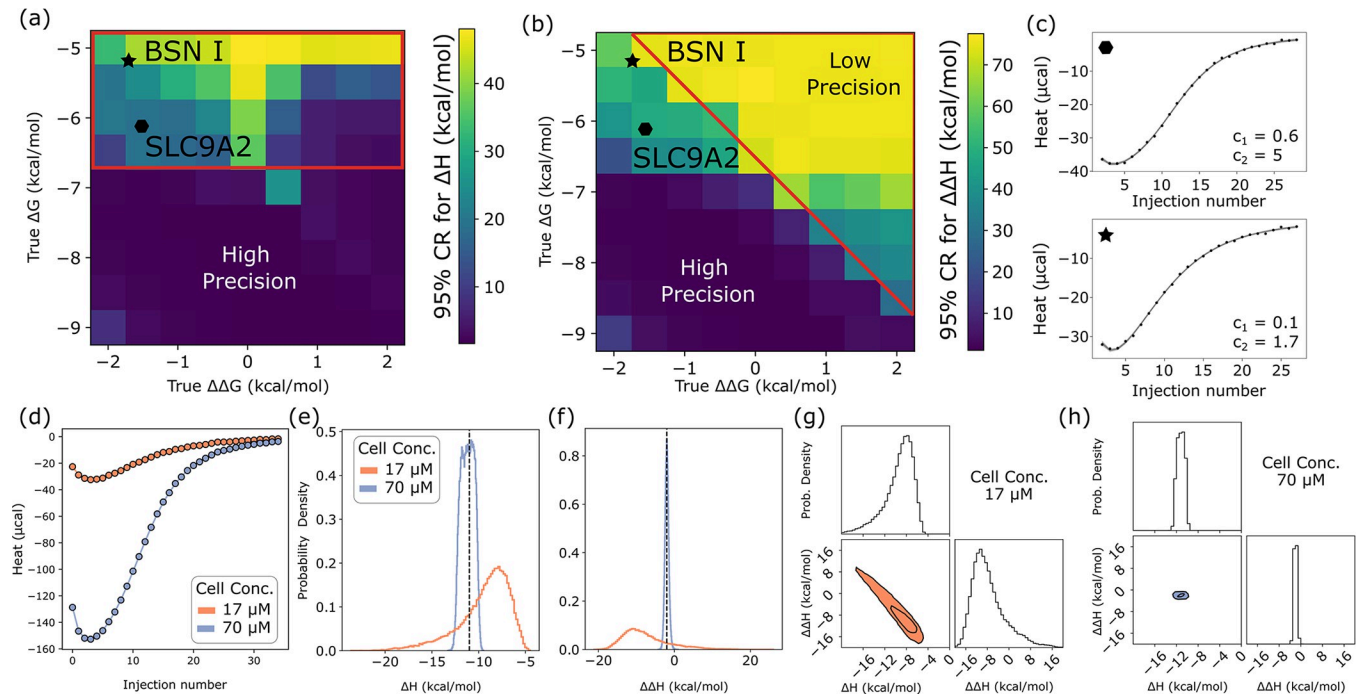


Fig 7. Strong dependence of posterior distributions on model parameters. (a,b) Graphical depiction of enthalpy uncertainty on a grid of ΔG and $\Delta\Delta G$ values, generated from the Bayesian posterior for each grid point based on synthetic data (with fixed $\Delta H = -10$, $\Delta\Delta H = -1.5$ kcal/mol). Boxes are colored by the width of the 95% credibility region for ΔH (a) and $\Delta\Delta H$ (b), with lighter colors corresponding to wider credibility regions (color bars). Red polygons demonstrate where each K_d (K_{d1} for left, K_{d2} for right) is greater than $17 \mu M$, which is the cell concentration set for these synthetic isotherms. Black symbols indicate mean values for experimental isotherms for binding for BSN motif I (star) and SLC9A2 (octagon), for comparison. (c) Isotherms for binding between LC8 and SLC9A2 (top) and BSN I (bottom). C values for each step of binding, based on mean values taken from the posterior distribution are shown. (d) Synthetic isotherms designed to mimic BSN I ($\Delta G = -5.1$, $\Delta\Delta G = -1.7$, $\Delta H = -11$, $\Delta\Delta H = -2$), simulated at cell concentrations of 17 (orange) and 70 (μM) (μM (syringe concentrations at 900 and $2000 \mu M$ respectively)). (e,f) One-dimensional probability distributions for binding enthalpy (e) and change in enthalpy (f) for the isotherms in panel (d). (g,h) Two-dimensional marginal distributions (plotting ΔH against $\Delta\Delta H$ for isotherms in panel (d)). Isotherms at $17 \mu M$ in (g) and $70 \mu M$ in (h). Dimension widths are fixed for both plots for better comparison.

<https://doi.org/10.1371/journal.pcbi.1011059.g007>

(Table 1), was correlated with increased uncertainty, i.e., broader posterior marginals, in binding parameters, especially ΔH and $\Delta\Delta H$. For this purpose, we created a series of synthetic isotherms on a grid of ΔG and $\Delta\Delta G$ values and determined posterior distributions for each isotherm. Two-dimensional heat maps of the width of these distributions across ΔG - $\Delta\Delta G$ space (Fig 7A and 7B) capture trends in our ability to determine model parameters.

Generally, we lose precision in binding enthalpy when binding is weaker, although there are nuances. Interestingly, the relationship appears to differ somewhat between ΔH (Fig 7A) and $\Delta\Delta H$ (Fig 7B). For ΔH , the primary dependence appears to be on the value of ΔG , with precision decreasing when $\Delta\Delta G$ is 0 or negative (top left corner of 7A). However, the precision for $\Delta\Delta H$ depends strongly on both ΔG and $\Delta\Delta G$, with the worst precision found in the top right quarter of the plot. This trend is largely consistent with a fundamental principle in calorimetric experimental design based on the experimental parameter c , defined as $c = n[\text{cell}]/K_d$ where n is the number of binding sites on the titrand, effectively a ratio between the cell concentration and the binding affinity. While the importance of c is debated [48], doctrine is that $5 < c < 500$ is required to determine binding affinities from an isotherm, as binding is either too weak or too strong for the isotherm to be information-rich outside of the 5–500 range of c .

Consistent with the preferred range for c , we see losses in precision as the values of c for our experiments decrease. Our cell concentration for synthetic isotherms is $17 \mu M$, meaning that c is ~ 1 at $\Delta G = -6.5$ kcal/mol. While c is primarily discussed only in the context of binding

models with a single affinity, with two-site binding we can calculate two separate c values (c_1 and c_2), for the first and second step of binding, respectively. Regions of the heat map with the lowest precision where $c_1 < 1$ are boxed in red in the ΔH heat map (Fig 7A), and $c_2 < 1$ is boxed in the $\Delta\Delta H$ heat map (Fig 7B). The presence of two c values complicates isotherm analysis, in some cases producing isotherms that appear tight-binding by visual inspection (e.g. SLC9A2, BSN I, Fig 7C), but have one or more values of c outside of the informative range, consistent with our observation that these isotherms have poorly determined binding enthalpy.

To examine whether increasing c would improve precision, we generated synthetic isotherms to mimic BSN I, at $\Delta G = -5.1$ and $\Delta\Delta G = -1.7$ kcal/mol, and applied the model at two cell concentrations, 17 and 70 μM (Fig 7D). Consistent with our expectations, posterior distributions for binding enthalpies narrowed dramatically (Fig 7E and 7F) at the 70 μM concentration, although it is worth noting that this improvement in precision is due partly to the increased signal-to-noise ratio at higher concentrations (S10 Fig). Based on these results, for isotherms like the SLC9A2 or BSN I titrations, performing additional experiments at increased analyte concentrations should resolve the poor precision in enthalpy determination. This provides an example of how microscopic binding parameters influence precision and demonstrates that future experiments designed to investigate multisite binding will benefit from consideration of how experimental conditions relate to the energy of each step of a multistep binding interaction.

Discussion

This work develops binding models for isothermal titration calorimetry (ITC) data of proteins with two symmetrical sites to extract thermodynamics parameters for each binding event and thus assess cooperativity. One such protein is the dimeric hub LC8, which binds over 100 client proteins at the same site. The essential role LC8 plays in regulating a variety of cell functions [5,9,49] motivates detailed mechanistic understanding of what drives recognition to its diverse partners. Using a Bayesian framework, we sought to determine precisely how much information can be extracted from a single ITC isotherm and examine how uncertainty in analyte concentration impacts model parameters, an investigation greatly aided by simulated ‘synthetic’ isotherms with known parameters. Building on prior work [32,36,37,39], we have advanced Bayesian analysis of binding, and applied it to rigorous biophysical characterization of LC8 dimer binding to short client peptides, as well as binding between the dimeric coiled-coil domain of NudE and the intermediate chain of dynein. We also used synthetic data to unambiguously separate effects of experimental error from system-intrinsic limitations imposed and define the limits of intrinsic (in)tractability in calorimetry.

Bayesian inference in binding analysis, leveraging synthetic data

“How much information is contained in an ITC isotherm?” is a fundamental biophysics question that Bayesian inference is uniquely suited to answer. Building on prior work [32,36,39], we have improved the ability of the Bayesian approach to account for the intrinsic uncertainty in *both* titrant and titrand concentrations. Our approach was motivated in large part by the recognition of a mathematical “degeneracy” in ITC analysis, i.e., the existence of multiple solutions even in the absence of experimental noise, which prevents inference of a fully unique set of thermodynamic parameters. This degeneracy holds for simple 1:1 binding and apparently for arbitrary stoichiometry.

While fitting ITC data to multisite binding and other complex models is challenging, Bayesian inference yields “posterior” joint probability distributions for model parameters, providing

a full description of parameter uncertainties and correlations consistent with any prior assumptions. Analogous investigation using traditional least-squares fitting commonly relies on post-analysis techniques such as the error-surface analysis implemented in SEDPHAT [21], which employs a series of maximum-likelihood fits that approximate the marginal probability distribution for a given model dimension being evaluated [40]. While such analyses may be adequate in many cases, they are not integrated into many fitting pipelines, including Origin, and only rarely employed. Bayesian inference offers a direct route to uncertainties and correlations between parameters without relying on a maximum-likelihood approximation [40]. The posterior distributions—the joint distribution over all binding parameters—fundamentally answer the question of the information contained in an isotherm [32,39].

Our work has benefited greatly from the use of synthetic isotherms. Built from known thermodynamic parameters, the value of synthetic isotherms as an aid in experimental design is well-established [21,32], and our Bayesian pipeline quantifies the relative information content of each generated isotherm. Synthetic isotherms have allowed us to test and troubleshoot our pipeline (Fig 3), probe the information content of isotherms under variable conditions of concentration and priors (Figs 3 and S5), and examine how thermodynamic parameters themselves impact our ability to determine information from isotherms, resulting in the heatmap of relative tractability (Fig 7). In the context of multi-isotherm modeling, utilizing synthetic data to design new experiments, similar to what is possible with fitting in the program SEDPHAT [21,35] will be particularly valuable.

Using synthetic data, our investigation of how concentrations impact model parameters has shown that uncertainty in concentration induces uncertainty in binding enthalpy, but has a reduced impact on free energy. This agrees well with results from Nguyen et al. [32] on 1:1 binding, indicating that the concentration-enthalpy relationship applies to all binding models. Thus, while the individual concentrations may be indeterminable from the model alone, the ratio of concentrations can be readily determined, provided the underlying stoichiometry of binding is known (S1 Table).

From a single experimental isotherm, we sample marginal posterior distributions with widths as narrow as 1–2 kcal/mol for a two-site model with four thermodynamic parameters. Uncertainties on this scale are consistent with other Bayesian analyses [36,39]. In addition to uncertainty due to experimental noise and correlative parameter relationships such as between ΔG and $\Delta\Delta G$, uncertainty arises from our ‘skeptical’ consideration of analyte concentrations (priors of $\pm 10\%$ for LC8, up to $\pm 50\%$ for peptides). While free energy and enthalpy parameters for individual binding steps cannot always be determined with precision, the total values accounting for both steps can be determined with less uncertainty (S1 Table and S3 Fig). Uncertainty can be reduced by careful concentration determination through multiple methods, and the use of global models derived from multiple isotherms at varied concentrations and concentration ratios.

Practical limitations of Bayesian sampling and global modeling

Bayesian statistical analysis relies on Markov chain Monte Carlo (MCMC) sampling, which requires simulating a sufficient number of steps to adequately explore the parameter space, potentially including a need to locate and sample multiple probability peaks (akin to energy basins in conformation space). For our ITC data analysis, simple MCMC sampling methods proved inadequate to sample the model space, even following sampling times of several days and over 4 million samples. While the ensemble sampler [50] used by us and others applying Bayesian models to ITC [36,39] has been robust for our purposes, adequate sampling continues to be an important limitation, especially when considering analysis of more complex models. For all work presented here, sampling ran at an average rate of around 7–8 iterations per

second on four cores, leading to wall-clock sampling times near two hours for models requiring 50,000 samples and 50 chains. This was readily feasible, even with replicates and varying conditions. More complex models could require significantly more sampling, although there is no simple scaling law that applies because of the difficult-to-predict nature of the parameter-space ‘landscape’. Global modeling of multiple isotherms may also require additional sampling: as isotherms are added to a global model, each one brings with it a new set of nuisance parameters (4 per isotherm in our work—see [Methods](#)). In our hands, two-isotherm global models could be sampled at a similar rate of 7–8 iter/s but required as many as 200,000 samples and therefore took closer to 8 hours, but nonetheless could be well-sampled within half a day. While global models of technical replicates may improve signal to noise ratios, ideally, global experiments should be designed with the intent of covering several experimental conditions [36,39], and all experiments must be high quality to ensure they contribute to global fits. Efficient sampling of global models is an ongoing research direction for us.

Cooperativity in two-site binding

Our data show that LC8 binds client proteins with positive cooperativity. Of the 7 peptides examined here, Bayesian analysis for all except one (BIM) yields a negative $\Delta\Delta G$ value, confirming our hypothesis that cooperativity drives LC8 binding [18] for these peptides. We additionally test binding between the intermediate chain (IC) of dynein and cargo adaptor NudE and find no evidence of cooperativity in the interaction. For the present study, we selected test LC8-binding isotherms with preference for two criteria we anticipated would leverage modeling: (1) tight-binding to LC8 and (2) an isotherm shape that breaks from a strict sigmoid. This selection process makes it difficult to say with certainty whether the cooperativity we see with LC8-client binding is unique to these clients or universal to LC8 binding. Nevertheless, we can draw some conclusions about the mechanism and function of cooperativity.

The mechanism of cooperativity in LC8-client binding appears to be entropically driven. While entropy is often the term with the widest distribution ([Table 1](#)), owing to its dependence on both the free energy and the enthalpy, there is a clear trend in our results towards positive $T\Delta\Delta S$ values, which equates to the second binding step being more entropically favorable than the first. Relatedly, NMR dynamics measurements indicate LC8’s flexible core is rigidified on binding to clients [17,51]. Since LC8-binding cooperativity necessarily requires some change in the structural ensemble of LC8, it is possible that the first binding step can be thought of as ‘paying up-front’ for the entropic cost of both binding steps—i.e., rigidifying the whole LC8 core. This mechanism would also allow for variation in cooperativity on a per-peptide basis, as the degree of rigidification in the core seen by NMR is dependent on client sequence [17]. Future molecular dynamics simulations can examine the differences in rigidity of the LC8 core in different bound states and across binding to different peptides.

While LC8-client complexes are varied, the putative functional unit of most LC8-client interactions is a 2:2 bound structure, where LC8 promotes dimerization in client proteins [52–54]. One possible underlying function of cooperative binding is therefore that it acts as a driver of the formation of the 2:2 bound state and suppresses the nonfunctional 1:2 intermediate. LC8 contrasts interestingly with NudE-IC binding in this respect. Recent work has demonstrated that pre-dimerization by other proteins is essential for the NudE-IC interaction when measured in context of the full length protein [45]. As such, there is no analogous functional role for cooperativity to play in IC binding. Prior work has proposed that positive cooperativity could drive the formation of homologous complexes [18], with the same client bound to each LC8 motif. This hypothesis suggests that binding at one site essentially promotes binding of the same client at the second site, through small adjustments along the LC8 dimer interface.

Future competition assays—titrations of a client into LC8 samples pre-bound with a different client—may help to answer this question.

New questions arise when considering LC8 clients with multiple binding motifs. Proteins such as the nucleoporin Nup159 or the transcription factor ASCIZ contain several LC8 motifs in succession and form large ladder-like complexes with LC8. Complexes between LC8 and multivalent clients are often highly heterogeneous in both stoichiometry and conformation [7,55,56], complicating their analysis. Particularly in the case of ASCIZ, the fully bound state is highly disfavored by negative cooperativity [7,56], thought to be mediated through the linker sequences between LC8 motif, which ensures that ASCIZ is sensitive to LC8 even at high LC8 concentrations. Investigating the thermodynamics of such complexes necessitates methods able to dissect the complicated network of binding interactions dependent both on the individual motifs as well as on the lengths and structures of linkers between motifs. This work represents a first step towards such investigations and lays a framework upon which analyses of multivalent LC8 interactions can be built.

Concluding remarks and future steps

Bayesian inference has allowed us to characterize multi-site binding cooperativity with high confidence for two different protein-protein interactions of 2:2 stoichiometry, despite uncertainties in analyte concentrations and inherent limitations of ITC. Our analysis was enabled by improvements to prior work [32,36,39] in treating concentration uncertainties, and further demonstrates the value of Bayesian inference to ITC analysis. We used synthetic data to systematically characterize the uncertainty landscape for 2:2 binding based on both intrinsic binding properties and experimental conditions, an approach that can readily be extended to other models.

We examined two dimeric systems, the hub protein LC8 and the coiled coil domain of NudE. For LC8, every client peptide studied showed evidence of cooperative binding, confirming hypotheses from a decade ago [18]. In contrast, the dynein NudE/IC complex showed minimal evidence of cooperativity, consistent with the fact that in biological settings, NudE binds to IC in a dimeric state, suggesting allostery would serve no purpose in the interaction. The ability to reliably characterize these interactions also serves as an important step toward quantitative characterization of multivalent LC8-multivalent client complexes, which, due to their complexity, evade straightforward investigation.

While our focus here has been on two-step symmetric-site binding systems, Bayesian methods can be applied to other complex models investigated by ITC. Measurement of complex multivalent systems, enthalpy-entropy compensation, and ternary complexes or competition binding are all likely to benefit from analysis under a Bayesian framework. Although there is a limit on how much information can be gained from individual isotherms, investigation utilizing synthetic data can guide design, to help determine experimental conditions that maximize gain from additional ITC experiments within a given system.

Methods

1:1 binding

For 1:1 binding, we used the quadratic model as described in Eqs 1 and 2 in the results above. The heat of each injection i , (dQ_i) was calculated using the following equation:

$$dQ_i = Q_i + \frac{V_i}{V_0} \left(\frac{Q_i - Q_{i-1}}{2} \right) - Q_{i-1} + \Delta H_0 \quad (5)$$

Where V_i is the volume of injection i . This binding model is identical to the model used in Origin's identical sites model when $n = 1$. ΔH_0 is a correction term to account for heat of dilution, buffer mismatch, and other effects that may apply a flat shift to binding heat.

Two-site binding

Two-site binding is modeled in a standard fashion, such as in the binding polynomial model [57] as:



Under this scheme, each binding affinity is as follows:

$$\frac{1}{2} K_{d1} = \frac{[X][M]}{[MX]} \quad (7)$$

$$2K_{d2} = \frac{[X][MX]}{[MX_2]} \quad (8)$$

where K_{d1} and K_{d2} are the affinities for the first and second microscopic, site-resolved binding steps. Factors of $\frac{1}{2}$ and 2 represent adjustments between the individual site-resolved affinities for either symmetric indistinguishable binding step and the 'macroscopic' binding constants (i.e. $K_{macro1} = \frac{1}{2} K_{d1}$). Here and throughout, we refer to binding affinities and energies by their microscopic value. The total concentrations of X and M can be written as:

$$[M_t] = [M] + [MX] + [MX_2] \quad (9)$$

$$[X_t] = [X] + [MX] + 2[MX_2] \quad (10)$$

Through rearrangement and substitution of Eqs 7 and 8, the total concentration equations can be rewritten only in terms of [M] and [X], the concentrations of free macromolecule and ligand:

$$[M_t] = [M] + 2 \frac{[X][M]}{K_{d1}} + \frac{[X]^2[M]}{K_{d1}K_{d2}} \quad (11)$$

$$[X_t] = [X] + 2 \frac{[X][M]}{K_{d1}} + 2 \frac{[X]^2[M]}{K_{d1}K_{d2}} \quad (12)$$

This system of equations is solved numerically for each given injection point to determine the unbound concentrations [M] and [X]. In our implementation, numerical solutions are calculated using the 'optimize.root' function in the scipy python library, using a Levenberg-Marquardt least-squares optimization. With both free concentrations determined, the system heat can be calculated:

$$\frac{Q}{V_0} = \Delta H_1[MX] + (\Delta H_1 + \Delta H_2)[MX_2] \quad (13)$$

where ΔH_1 and ΔH_2 are the enthalpies of binding step one and two respectively. The concentrations of each bound state can be calculated from [X] and [M] and Eqs 7 and 8. As in the 1:1 binding model, Eq 5 is used to calculate the observed heat of injection, dQ_i , for each injection.

Degeneracy in two-step binding

When protein concentrations are included as model parameters, degenerate solutions are introduced. As outlined in the manuscript, the degeneracy is exposed from the following transformation:

$$\begin{aligned}
 [M_t] &\rightarrow \alpha[M_t] \\
 [X_t] &\rightarrow \alpha[X_t] \\
 K_{d1} &\rightarrow \alpha K_{d1} \\
 \Delta G_1 &\rightarrow \Delta G_1 + RT \log \alpha \\
 K_{d2} &\rightarrow \alpha K_{d2} \\
 \Delta G_2 &\rightarrow \Delta G_2 + RT \log \alpha \\
 \Delta H_1 &\rightarrow \frac{\Delta H_1}{\alpha} \\
 \Delta H_2 &\rightarrow \frac{\Delta H_2}{\alpha}
 \end{aligned} \tag{14}$$

Here, α can be any positive number. Following this transformation, the equations used to calculate $[X]$ and $[M]$ (Eqs 7 and 8) are transformed:

$$\alpha[M_t] = [M] + 2 \frac{[X][M]}{\alpha K_{d1}} + \frac{[X]^2[M]}{\alpha^2 K_{d1} K_{d2}} \tag{15}$$

$$\alpha[X_t] = [X] + 2 \frac{[X][M]}{\alpha K_{d1}} + 2 \frac{[X]^2[M]}{\alpha^2 K_{d1} K_{d2}} \tag{16}$$

In these transformed concentration-sum equations, the new solutions for both $[X]$ and $[M]$ are exactly the previous solutions multiplied by α , as can be verified by substitution. Finally, applying the transformed values into the equation for Q yields

$$\frac{Q}{V_0} = 2 \frac{\Delta H_1}{\alpha} \frac{\alpha[X] \alpha[M]}{\alpha K_{d1}} + \frac{(\Delta H_1 + \Delta H_2)}{\alpha} \frac{\alpha[M] (\alpha[X])^2}{\alpha^2 K_{d1} K_{d2}} \tag{17}$$

As in the 1:1 binding model, cancellation of α shows there is no change in the value of Q for any α value. This demonstrates the degeneracy for 2:2 binding, which we can expect to generalize to higher stoichiometries.

Bayesian inference

Bayesian inference is a method to calculate a “posterior” *distribution* of model parameter values based on prior assumptions (encoded as prior distributions for parameters presumed to hold in the absence of data) and the data. In general, as more data is analyzed, the influence of the prior will decrease [58,59]. The posterior distribution of parameters provides rich information such as the parameter means and confidence intervals (technically “credibility regions”), in addition to correlation information regarding whether and how parameters vary together.

Bayesian inference is based on Bayes' rule [58,60] which enables us to infer a distribution of parameters θ (e.g., binding free energy and enthalpy, etc.) consistent with a given set of data D (e.g., ITC isotherms):

$$P(\theta|D) = P(D|\theta)P(\theta)/P(D) \quad (18)$$

where $P(\theta|D)$ is the (posterior) probability distribution of the model parameters, θ , given the data, D ; $P(D|\theta)$ (the likelihood) is the probability distribution of the data given the model parameters and is given below; $P(\theta)$ (the prior) is the probability of the model parameters, specified below; and $P(D)$ (the evidence) is the probability of the data. For a given set of data, the unknown denominator $P(D)$ is constant, independent of parameters, so it does not affect the inference of posteriors. Typically, it is not possible to analytically solve Bayes' rule, so numerical methods such as Markov chain Monte Carlo are used to determine the target (posterior) distribution [61–63]. Details of our implementation are given below.

Bayesian model

Following prior work [30,39], we assume the data has Gaussian noise with a mean of zero and an unknown standard deviation. The ITC model parameters θ include concentration terms ($X_{\text{initial}}, M_{\text{initial}}$) and thermodynamic terms ($\Delta G, \Delta\Delta G, \Delta H, \Delta\Delta H$), as well as the nuisance parameters (ΔH_0 and σ) for heat of dilution and Gaussian noise. We use uniform prior distributions for the model parameters specified below and the unknown noise standard deviation unless otherwise stated. For global models (e.g. Figs S8 and 6), while it may be possible to assume a global noise or concentration model, we instead elected to apply global models with an additional set of concentration and nuisance parameters for each additional isotherm (bringing the total parameter count up to 12 for two-isotherm models). Uniform prior ranges for thermodynamic parameters were identical for all models, listed in S4 Table. For nuisance parameters ΔH_0 and σ , uniform priors of -10 to 10 μcal and 0.001 to 1 μcal respectively were used in all models.

The likelihood for a set of data $D = \{x_1, x_2, \dots\}$, denoted ($p(D|\theta)$), is the product of the probabilities at all data points x_i based on a normal distribution of standard deviation σ centered around $\mu_i(\theta)$, the calculated value of point i for the binding model and parameters θ . It therefore takes the following form:

$$p(D|\theta) = \prod_i \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x_i - \mu_i)^2}{2\sigma^2}\right\} \quad (19)$$

and we note that σ is assumed unknown and sampled as part of the Bayesian inference process. When the priors are uniform, as we most often assume, the posterior is simply proportional to the likelihood given here.

Sampling

We use an affine-invariant Markov chain Monte Carlo sampling method [64] to perform Bayesian inference, as also used by Duvvuri et al. [39] and Cardoso et al. [36]. The affine-invariant sampler is an ensemble-based method in which multiple walkers move through the sample space in a correlated fashion. We empirically found this method to sample significantly better than the standard Metropolis-Hastings [61,62] sampler for our model. In our hands, the Metropolis-Hastings method was unable to converge on the target distribution after 4,000,000 sampling steps, whereas the affine-invariant sampler was able to converge after 100,000 sampling steps.

Implementation

We used the EMCEE package [50] in Python to perform the affine sampling, using a 20%:80% mix of the “differential evolution” and “stretch” move sets with 25–50 walkers. For each experiment, 3 replicas are run for 50,000–200,000 sampling steps/replica until convergence. Each replica converged, as determined by the autocorrelation time, where sampled steps must be greater than 50x the autocorrelation. Convergence was additionally assessed through examination of posterior distributions from model replicas, which were nearly identical in all cases (S6 Fig). This implementation runs at ~8 samples for each walker per second on 4 cores of a node on the Oregon State College of Science computing cluster, leading to sampling times of around 1 hour and 40 minutes for models requiring only 50,000 samples and up to 7 hours for models requiring 200,000 samples. Additional discussion of the sampling method and its implementation can be found in our best practices supplement (S1 Document).

The code, data, and an example notebook are available at: https://github.com/ZuckermanLab/Bayesian_ITC

Synthetic isotherms

Synthetic isotherms for 1:1 and two-site binding were generated following Eq 2 for 1:1 binding and Eqs 11, 12, and 10 for two-site binding. Parameters were chosen to mimic typical experimental conditions employed in our group. For 1:1 binding (S5 Fig), we used ΔG and ΔH values of -8 and -12 kcal/mol respectively, and concentrations of 34 μM in the cell and 500 μM in the syringe. For two-site binding, varied thermodynamic parameters were used (e.g. Figs 3 and 7), but concentrations were fixed at 17 μM in the cell and 500 μM in the syringe. Synthetic isotherms used a cell volume of 1.42 mL and a temperature of 25 C. For synthetic isotherms, we simulated one injection of 2 μL followed by 34 injections of either 6 μL (Isotherms in Figs 3, S1, S4 and S5) or 10 μL (BSN mimic isotherms, Figs 7D–7H and S10). All isotherms were calculated with a ΔH_0 of 0 μcal , and added synthetic noise from a Gaussian distribution with a mean of 0 and standard deviation 0.2 μcal (except the high noise BSN I isotherm, S10 Fig). To accurately replicate experimental conditions, we eliminated the first injection when applying models to this data.

Validation of model credibility regions

To validate the model’s credibility regions, we generated replicate isotherms each with new random noise, and we sampled posterior distributions for each (S2 Fig). For the full model including concentrations we used 100 replicates, and for the reduced model without concentrations we used 50. Isotherms were generated from the experimental parameters used in the isotherm in Fig 3B ($\Delta G = -7$, $\Delta H = -10$, $\Delta\Delta G = -1$, $\Delta\Delta H = -1.5$ kcal/mol) at fixed concentrations of 17 μM in the cell and 500 μM in the syringe, standard deviation of 0.2 μcal for injection heat. For the full model, to simulate experimental conditions, we randomly selected ‘measured’ concentrations from the $\pm 10\%$ uniform prior. Uncertainties in the credibility values presented in S2 Fig are standard deviations of 1000 bootstrapped samples.

Supporting information

S1 Fig. MCMC traces and marginal distributions for all model parameters for a synthetic isotherm. (a) MCMC traces consisting of 50 chains for all model parameters for a synthetic isotherm (Fig 3B). Chains are thinned by a factor of 50 for visibility. (b) One and two-dimensional marginal distributions for all model parameters for a synthetic isotherm (Fig 3B), with contours in the two-dimensional plots set at 95 (yellow), 75 (orange), 50 (purple) and 25%

(black) confidence. Red lines and dots indicate true values for the synthetic isotherm. Correlations between some parameters are apparent, including between concentrations, as well as between each concentration and the thermodynamic parameters, especially ΔH .

(PDF)

S2 Fig. Validation of uncertainty in credibility regions. Plots of predicted versus observed credibility for each thermodynamic parameter and analyte concentrations for our test synthetic isotherm conditions (Figs 3B and S1). Black circles are the complete model including concentrations, while red circles are from a simplified model with concentrations removed. Error bars are standard deviations across 1000 bootstrapped samples.

(PDF)

S3 Fig. Distributions of thermodynamic parameters plotted with total free energies and enthalpies. Each plot shows a set of either ΔG and $\Delta\Delta G$ or ΔH and $\Delta\Delta H$, along with the ‘total’ value for that parameter, i.e. $2\Delta G + \Delta\Delta G$ or $2\Delta H + \Delta\Delta H$. The distributions for this sum value are often narrower than the individual parameters, as the total enthalpy and free energy of binding can be determined with higher precision from a given isotherm than the individual values. ΔG , ΔH are the energy and enthalpy of binding step 1, while $\Delta G + \Delta\Delta G$, $\Delta H + \Delta\Delta H$ are the energy and enthalpy of binding step 2, making the total values reported here the energy and enthalpy of both binding steps combined.

(PDF)

S4 Fig. Marginal distributions comparing models with uniform and normal-distribution priors. Distributions are taken from models on an identical synthetic isotherm generated from parameters $\Delta G = -7$, $\Delta\Delta G = -1$, $\Delta H = -10$, $\Delta\Delta H = -1.5$, $[\text{peptide}]_{\text{initial}} = 500$, $[\text{LC8}]_{\text{initial}} = 17$, $\Delta H_0 = 0$ and $\sigma = 0.2$. All priors are identical except for $[\text{peptide}]_{\text{initial}}$, where the uniform prior model (red) was run with a $\pm 10\%$ of stated value uniform prior, and the normal prior model (black) was run with a normal distribution prior with standard deviation = 1% of stated value.

(PDF)

S5 Fig. Effect of concentration priors on marginal posterior distributions for thermodynamic parameters in a 1:1 binding model. Distributions are taken from models on an identical synthetic isotherm generated from parameters $\Delta G = -8$, $\Delta H = -12$, $[X]_{\text{initial}} = 500$, $[M]_{\text{initial}} = 34$, $\Delta H_0 = 0$ and $\sigma = 0.2$. Model priors are uniform distributions of varied width in each plot for $[X]_{\text{initial}}$ and $[M]_{\text{initial}}$, varied from $\pm 1\%$ to $\pm 50\%$.

(PDF)

S6 Fig. Example marginal distributions of replicate models for the LC8-SPAG5 interaction. Each model replicate is run on an identical isotherm with a different random seed dictating random starts for MCMC chains and trial move selections. Each model returns near-identical marginal distributions.

(PDF)

S7 Fig. two dimensional marginal distributions of enthalpy for selected isotherms. Marginal distributions for BSN I, SLC9A2, and GLCCI are shown, each of which has wide 1D distributions for both ΔH and $\Delta\Delta H$. Enthalpy parameters are closely correlated, resulting in a diagonal two-dimensional distribution within the enthalpy space.

(PDF)

S8 Fig. Marginal distributions for thermodynamic parameters for individual and global models for three LC8-peptide interactions. Distributions for each individual isotherm and

distributions for the global model are shown in purple, orange and green respectively. While the global model improves precision in determined parameters in some cases (e.g. GLCCI), in others it appears to follow the shape of the distributions for individual isotherms (e.g. BSN I). (PDF)

S9 Fig. Marginal distributions for thermodynamic parameters for IC-NudE binding isotherms. Distributions for each individual isotherm and distributions for the global model are shown in green, orange, and purple respectively. (PDF)

S10 Fig. Marginal distributions for thermodynamic parameters for BSN-like synthetic isotherms. (a) one-dimensional distributions for thermodynamic parameters for synthetic isotherms ($\Delta G = -5.1$, $\Delta\Delta G = -1.7$, $\Delta H = -11$, $\Delta\Delta H = -2$) generated with cell concentrations of 17 (orange) and 70 μM (blue and green). Syringe concentrations are 900 μM and 2000 μM respectively. The orange and blue isotherms are generated with noise taken from a gaussian distribution of width $\sigma = 0.2$ μcal , while the green is generated with $\sigma = 0.8$ ucl . (b) two-dimensional marginal distributions for the same models as (a), in the ΔH - $\Delta\Delta H$ dimension. Contours are drawn at 95 and 50% probability density. While raising the synthetic experimental concentration dramatically improves precision in all model parameters, much of this is due to the increased S/N ratio associated with the higher concentration. Scaling synthetic model noise with the increase in cell concentration reduces the precision of model enthalpies, although they are still narrower than the distributions for the low concentration isotherm. (PDF)

S1 Table. Credibility regions for ‘sum’ thermodynamic parameters and ratios of concentrations. 95% credibility region from sampled posterior distributions for the ΔG sum ($2\Delta G + \Delta\Delta G$) and ΔH sum ($2\Delta H + \Delta\Delta H$) as well as the ratio of concentrations ($[\text{peptide}]/[\text{LC8}]$). Credibility regions for ΔG and ΔH sums are frequently narrower than the credibility regions for individual parameters (Table 1). (PDF)

S2 Table. Ranges for thermodynamic parameters for LC8-client binding when modeled with $\pm 20\%$ LC8 concentration. Table 1 in the main text contains equivalent information at $\pm 10\%$ LC8 concentration. Values delineate 95% Bayesian credibility regions from sampled posterior distributions, when modeled with $\pm 20\%$ priors for LC8 concentration. Distributions are largely very similar to those presented in Table 1, with a slight decrease in precision. BSN I, for which posterior distributions are significantly broader, is the only notable exception. (PDF)

S3 Table. Binding parameters determined from identical sites model fits, as published in Jespersen et. al. (2019) [9]. (PDF)

S4 Table. Model priors and sampling lengths for all isotherms. (PDF)

S1 Document. Best practices for the application of Bayesian statistical models to isothermal titration calorimetry. (PDF)

Acknowledgments

We appreciate helpful discussions with John Chodera, David Minh, and Trung Hai Nguyen.

Author Contributions

Conceptualization: Aidan B. Estelle, August George, Elisar J. Barbar, Daniel M. Zuckerman.

Data curation: Aidan B. Estelle.

Formal analysis: Aidan B. Estelle.

Funding acquisition: Elisar J. Barbar, Daniel M. Zuckerman.

Investigation: Aidan B. Estelle.

Methodology: Aidan B. Estelle, August George, Daniel M. Zuckerman.

Project administration: Elisar J. Barbar, Daniel M. Zuckerman.

Software: Aidan B. Estelle, August George.

Supervision: Elisar J. Barbar.

Validation: Aidan B. Estelle.

Visualization: Aidan B. Estelle.

Writing – original draft: Aidan B. Estelle.

Writing – review & editing: August George, Elisar J. Barbar, Daniel M. Zuckerman.

References

1. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Molecular Biology of the Cell. 4th ed. Garland Science; 2002.
2. Schneider R, Blackledge M, Jensen MR. Elucidating binding mechanisms and dynamics of intrinsically disordered protein complexes using NMR spectroscopy. *Curr Opin Struct Biol.* 2019; 54: 10–18. <https://doi.org/10.1016/j.sbi.2018.09.007> PMID: 30316104
3. Weng J, Wang W. Dynamic multivalent interactions of intrinsically disordered proteins. *Curr Opin Struct Biol.* 2020; 62: 9–13. <https://doi.org/10.1016/j.sbi.2019.11.001> PMID: 31783325
4. Barbar E. Dynein Light Chain LC8 Is a Dimerization Hub Essential in Diverse Protein Networks. *Biochemistry.* 2008; 47: 503–508. <https://doi.org/10.1021/bi701995m> PMID: 18092820
5. Rapali P, Szenes Á, Radnai L, Bakos A, Pál G, Nyitray L. DYNLL/LC8: a light chain subunit of the dynein motor complex and beyond. *FEBS J.* 2011; 278: 2980–2996. <https://doi.org/10.1111/j.1742-4658.2011.08254.x> PMID: 21777386
6. Clark S, Nyarko A, Löhr F, Karplus PA, Barbar E. The Anchored Flexibility Model in LC8 Motif Recognition: Insights from the Chica Complex. *Biochemistry.* 2016; 55: 199–209. <https://doi.org/10.1021/acs.biochem.5b01099> PMID: 26652654
7. Clark S, Myers JB, King A, Fiala R, Novacek J, Pearce G, et al. Multivalency regulates activity in an intrinsically disordered transcription factor. Kay LE, editor. *eLife.* 2018; 7: e36258. <https://doi.org/10.7554/eLife.36258> PMID: 29714690
8. Clark SA, Jespersen N, Woodward C, Barbar E. Multivalent IDP assemblies: Unique properties of LC8-associated, IDP duplex scaffolds. *FEBS Lett.* 2015; 589: 2543–2551. <https://doi.org/10.1016/j.febslet.2015.07.032> PMID: 26226419
9. Jespersen N, Estelle A, Waugh N, Davey NE, Blikstad C, Ammon Y-C, et al. Systematic identification of recognition motifs for the hub protein LC8. *Life Sci Alliance.* 2019; 2: e201900366. <https://doi.org/10.26508/lsa.201900366> PMID: 31266884
10. Nyarko A, Song Y, Nováček J, Židek L, Barbar E. Multiple Recognition Motifs in Nucleoporin Nup159 Provide a Stable and Rigid Nup159-Dyn2 Assembly. *J Biol Chem.* 2013; 288: 2614–2622. <https://doi.org/10.1074/jbc.M112.432831> PMID: 23223634
11. Erdős G, Szaniszló T, Pajkos M, Hajdu-Soltész B, Kiss B, Pál G, et al. Novel linear motif filtering protocol reveals the role of the LC8 dynein light chain in the Hippo pathway. *PLOS Comput Biol.* 2017; 13: e1005885. <https://doi.org/10.1371/journal.pcbi.1005885> PMID: 29240760
12. West KL, Kelliher JL, Xu Z, An L, Reed MR, Eoff RL, et al. LC8/DYNLL1 is a 53BP1 effector and regulates checkpoint activation. *Nucleic Acids Res.* 2019; 47: 6236–6249. <https://doi.org/10.1093/nar/gkz263> PMID: 30982887

13. Jespersen NE, Leyrat C, Gérard FC, Bourhis J-M, Blondel D, Jamin M, et al. The LC8-RavP ensemble Structure Evinces A Role for LC8 in Regulating Lyssavirus Polymerase Functionality. *J Mol Biol.* 2019; 431: 4959–4977. <https://doi.org/10.1016/j.jmb.2019.10.011> PMID: 31634467
14. Luthra P, Jordan DS, Leung DW, Amarasinghe GK, Basler CF. Ebola Virus VP35 Interaction with Dynein LC8 Regulates Viral RNA Synthesis. *J Virol.* 2015; 89: 5148–5153. <https://doi.org/10.1128/JVI.03652-14> PMID: 25741013
15. Rodríguez Galván J, Donner B, Veseley CH, Reardon P, Forsythe HM, Howe J, et al. Human Parainfluenza Virus 3 Phosphoprotein Is a Tetramer and Shares Structural and Interaction Features with Ebola Phosphoprotein VP35. *Biomolecules.* 2021; 11: 1603. <https://doi.org/10.3390/biom11111603> PMID: 34827601
16. Hall J, Karplus PA, Barbar E. Multivalency in the Assembly of Intrinsically Disordered Dynein Intermediate Chain. *J Biol Chem.* 2009; 284: 33115–33121. <https://doi.org/10.1074/jbc.M109.048587> PMID: 19759397
17. Nyarko A, Hall J, Hall A, Hare M, Kremerskothen J, Barbar E. Conformational dynamics promote binding diversity of dynein light chain LC8. *Biophys Chem.* 2011; 159: 41–47. <https://doi.org/10.1016/j.bpc.2011.05.001> PMID: 21621319
18. Benison G, Karplus PA, Barbar E. The Interplay of Ligand Binding and Quaternary Structure in the Diverse Interactions of Dynein Light Chain LC8. *J Mol Biol.* 2008; 384: 954–966. <https://doi.org/10.1016/j.jmb.2008.09.083> PMID: 18948118
19. MicroCal. Data Analysis in Origin. 1998.
20. Freiburger L, Auclair K, Mittermaier A. Global ITC fitting methods in studies of protein allostery. *Methods.* 2015; 76: 149–161. <https://doi.org/10.1016/j.ymeth.2014.12.018> PMID: 25573261
21. Zhao H, Piszczek G, Schuck P. SEDPHAT—a platform for global ITC analysis and global multi-method analysis of molecular interactions. *Methods San Diego Calif.* 2015; 76: 137–148. <https://doi.org/10.1016/j.ymeth.2014.11.012> PMID: 25477226
22. Feng C, Roy A, Post CB. Entropic allostery dominates the phosphorylation-dependent regulation of Syk tyrosine kinase release from immunoreceptor tyrosine-based activation motifs. *Protein Sci.* 2018; 27: 1780–1796. <https://doi.org/10.1002/pro.3489> PMID: 30051939
23. Lee AL, Sapienza PJ. Thermodynamic and NMR Assessment of Ligand Cooperativity and Intersubunit Communication in Symmetric Dimers: Application to Thymidylate Synthase. *Front Mol Biosci.* 2018; 5. Available: <https://www.frontiersin.org/article/10.3389/fmolb.2018.00047> <https://doi.org/10.3389/fmolb.2018.00047> PMID: 29888227
24. Chodera JD, Mobley DL. Entropy-enthalpy compensation: Role and ramifications in biomolecular ligand recognition and design. *Annu Rev Biophys.* 2013; 42: 121–142. <https://doi.org/10.1146/annurev-biophys-083012-130318> PMID: 23654303
25. Czodrowski P, Sotriffer CA, Klebe G. Protonation Changes upon Ligand Binding to Trypsin and Thrombin: Structural Interpretation Based on pKa Calculations and ITC Experiments. *J Mol Biol.* 2007; 367: 1347–1356. <https://doi.org/10.1016/j.jmb.2007.01.022> PMID: 17316681
26. Steuber H, Czodrowski P, Sotriffer CA, Klebe G. Tracing Changes in Protonation: A Prerequisite to Factorize Thermodynamic Data of Inhibitor Binding to Aldose Reductase. *J Mol Biol.* 2007; 373: 1305–1320. <https://doi.org/10.1016/j.jmb.2007.08.063> PMID: 17905306
27. Leavitt S, Freire E. Direct measurement of protein binding energetics by isothermal titration calorimetry. *Curr Opin Struct Biol.* 2001; 11: 560–566. [https://doi.org/10.1016/s0959-440x\(00\)00248-7](https://doi.org/10.1016/s0959-440x(00)00248-7) PMID: 11785756
28. Velazquez-Campoy A, Kiso Y, Freire E. The Binding Energetics of First- and Second-Generation HIV-1 Protease Inhibitors: Implications for Drug Design. *Arch Biochem Biophys.* 2001; 390: 169–175. <https://doi.org/10.1006/abbi.2001.2333> PMID: 11396919
29. Tellinghuisen J, Chodera JD. Systematic errors in isothermal titration calorimetry: Concentrations and baselines. *Anal Biochem.* 2011; 414: 297–299. <https://doi.org/10.1016/j.ab.2011.03.024> PMID: 21443854
30. Contreras-Martos S, Nguyen HH, Nguyen PN, Hristozova N, Macossay-Castillo M, Kovacs D, et al. Quantification of Intrinsically Disordered Proteins: A Problem Not Fully Appreciated. *Front Mol Biosci.* 2018; 5. Available: <https://www.frontiersin.org/article/10.3389/fmolb.2018.00083> <https://doi.org/10.3389/fmolb.2018.00083> PMID: 30234128
31. Myszkowski DG, Abdiche YN, Arisaka F, Byron O, Eisenstein E, Hensley P, et al. The ABRF-MIRG'02 Study: Assembly State, Thermodynamic, and Kinetic Analysis of an Enzyme/Inhibitor Interaction. *J Biomol Tech JBT.* 2003; 14: 247–269. PMID: 14715884

32. Nguyen TH, Rustenburg AS, Krimmer SG, Zhang H, Clark JD, Novick PA, et al. Bayesian analysis of isothermal titration calorimetry for binding thermodynamics. *PLOS ONE*. 2018; 13: e0203224. <https://doi.org/10.1371/journal.pone.0203224> PMID: 30212471
33. Boyce SE, Tellinghuisen J, Chodera JD. Avoiding accuracy-limiting pitfalls in the study of protein-ligand interactions with isothermal titration calorimetry. *bioRxiv*; 2015. p. 023796. <https://doi.org/10.1101/023796>
34. MicroCal. VP-ITC Users Manual. 2001.
35. Houtman JCD, Brown PH, Bowden B, Yamaguchi H, Appella E, Samelson LE, et al. Studying multisite binary and ternary protein interactions by global analysis of isothermal titration calorimetry data in SED-PHAT: Application to adaptor protein complexes in cell signaling. *Protein Sci*. 2007; 16: 30–42. <https://doi.org/10.1110/ps.062558507> PMID: 17192587
36. Cardoso MVC, Rivera JD, Vitale PAM, Degenhardt MFS, Abiko LA, Oliveira CLP, et al. CALX-CBD1 Ca²⁺-Binding Cooperativity Studied by NMR Spectroscopy and ITC with Bayesian Statistics. *Biophys J*. 2020; 119: 337–348. <https://doi.org/10.1016/j.bpj.2020.05.031> PMID: 32574558
37. Nguyen TH, La VNT, Burke K, Minh DDL. Bayesian regression and model selection for isothermal titration calorimetry with enantiomeric mixtures. *PLOS ONE*. 2022; 17: e0273656. <https://doi.org/10.1371/journal.pone.0273656> PMID: 36173969
38. Minh DDL, Makowski L. Wide-Angle X-Ray Solution Scattering for Protein-Ligand Binding: Multivariate Curve Resolution with Bayesian Confidence Intervals. *Biophys J*. 2013; 104: 873–883. <https://doi.org/10.1016/j.bpj.2012.12.019> PMID: 23442966
39. Duvvuri H, Wheeler LC, Harms MJ. pytc: Open-Source Python Software for Global Analyses of Isothermal Titration Calorimetry Data. *Biochemistry*. 2018; 57: 2578–2583. <https://doi.org/10.1021/acs.biochem.7b01264> PMID: 29620867
40. Bevington PR, Robinson DK. *Data Reduction and Error Analysis for the Physical Sciences*. McGraw-Hill; 2003.
41. Nyarko A, Song Y, Barbar E. Intrinsic Disorder in Dynein Intermediate Chain Modulates Its Interactions with NudE and Dynactin. *J Biol Chem*. 2012; 287: 24884–24893. <https://doi.org/10.1074/jbc.M112.376038> PMID: 22669947
42. Wiseman T, Williston S, Brandts JF, Lin L-N. Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal Biochem*. 1989; 179: 131–137. [https://doi.org/10.1016/0003-2697\(89\)90213-3](https://doi.org/10.1016/0003-2697(89)90213-3) PMID: 2757186
43. Benison G, Chiodo M, Karplus PA, Barbar E. Structural, Thermodynamic, and Kinetic Effects of a Phosphomimetic Mutation in Dynein Light Chain LC8. *Biochemistry*. 2009; 48: 11381–11389. <https://doi.org/10.1021/bi901589w> PMID: 19863079
44. Anthis NJ, Clore GM. Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm. *Protein Sci Publ Protein Soc*. 2013; 22: 851–858. <https://doi.org/10.1002/pro.2253> PMID: 23526461
45. Jara KA, Loening NM, Reardon PN, Yu Z, Woonnimani P, Brooks C, et al. Multivalency, autoinhibition, and protein disorder in the regulation of interactions of dynein intermediate chain with dynactin and the nuclear distribution protein. Kay LE, Dötsch V, editors. *eLife*. 2022; 11: e80217. <https://doi.org/10.7554/eLife.80217> PMID: 36416224
46. Jie J, Löhr F, Barbar E. Interactions of Yeast Dynein with Dynein Light Chain and Dynactin: GENERAL IMPLICATIONS FOR INTRINSICALLY DISORDERED DUPLEX SCAFFOLDS IN MULTIPROTEIN ASSEMBLIES. *J Biol Chem*. 2015; 290: 23863–23874. <https://doi.org/10.1074/jbc.M115.649715> PMID: 26253171
47. Jie J. *Protein Disorder in the Evolution of Dynein Regulation*. Oregon State University; 2016.
48. Turnbull WB, Daranas AH. On the Value of c: Can Low Affinity Systems Be Studied by Isothermal Titration Calorimetry? *J Am Chem Soc*. 2003; 125: 14859–14866. <https://doi.org/10.1021/ja036166s> PMID: 14640663
49. Jespersen N, Barbar E. Emerging Features of Linear Motif-Binding Hub Proteins. *Trends Biochem Sci*. 2020; 45: 375–384. <https://doi.org/10.1016/j.tibs.2020.01.004> PMID: 32311332
50. Foreman-Mackey D, Hogg DW, Lang D, Goodman J. emcee: The MCMC Hammer. *Publ Astron Soc Pac*. 2013; 125: 306–312. <https://doi.org/10.1086/670067>
51. Hall J, Hall A, Pursifull N, Barbar E. Differences in Dynamic Structure of LC8 Monomer, Dimer, and Dimer–Peptide Complexes. *Biochemistry*. 2008; 47: 11940–11952. <https://doi.org/10.1021/bi801093k> PMID: 18942858
52. Becker JR, Cuella-Martin R, Barazas M, Liu R, Oliveira C, Oliver AW, et al. The ASCIZ-DYNLL1 axis promotes 53BP1-dependent non-homologous end joining and PARP inhibitor sensitivity. *Nat Commun*. 2018; 9: 5406. <https://doi.org/10.1038/s41467-018-07855-x> PMID: 30559443

53. Manneville J-B, Jehanno M, Etienne-Manneville S. Dlg1 binds GKAP to control dynein association with microtubules, centrosome positioning, and cell polarity. *J Cell Biol.* 2010; 191: 585–598. <https://doi.org/10.1083/jcb.201002151> PMID: 21041448
54. Wang L, Hare M, Hays TS, Barbar E. Dynein Light Chain LC8 Promotes Assembly of the Coiled-Coil Domain of Swallow Protein. *Biochemistry.* 2004; 43: 4611–4620. <https://doi.org/10.1021/bi036328x> PMID: 15078108
55. Mostofian B, McFarland R, Estelle A, Howe J, Barbar E, Reichow SL, et al. Continuum dynamics and statistical correction of compositional heterogeneity in multivalent IDP oligomers resolved by single-particle EM. *J Mol Biol.* 2022; 434: 167520. <https://doi.org/10.1016/j.jmb.2022.167520> PMID: 35245498
56. Reardon PN, Jara KA, Rolland AD, Smith DA, Hoang HTM, Prell JS, et al. The dynein light chain 8 (LC8) binds predominantly “in-register” to a multivalent intrinsically disordered partner. *J Biol Chem.* 2020; 295: 4912–4922. <https://doi.org/10.1074/jbc.RA119.011653> PMID: 32139510
57. Freire E, Schön A, Velazquez-Campoy A. Chapter 5 Isothermal Titration Calorimetry: General Formalism Using Binding Polynomials. *Methods in Enzymology.* Academic Press; 2009. pp. 127–155. [https://doi.org/10.1016/S0076-6879\(08\)04205-5](https://doi.org/10.1016/S0076-6879(08)04205-5) PMID: 19289205
58. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis Third edition.*: 677.
59. Hogg DW, Foreman-Mackey D. *Data Analysis Recipes: Using Markov Chain Monte Carlo.* *Astrophys J Suppl Ser.* 2018; 236: 11. <https://doi.org/10.3847/1538-4365/aab76e>
60. Bayes T. An essay towards solving a problem in the doctrine of chances. *Phil Trans R Soc Lond.* 1763; 53: 370–418.
61. Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika.* 1970; 57: 97–109. <https://doi.org/10.1093/biomet/57.1.97>
62. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of State Calculations by Fast Computing Machines. *J Chem Phys.* 1953; 21: 1087–1092. <https://doi.org/10.1063/1.1699114>
63. Sharma S. Markov Chain Monte Carlo Methods for Bayesian Data Analysis in Astronomy. *Annu Rev Astron Astrophys.* 2017; 55: 213–259. <https://doi.org/10.1146/annurev-astro-082214-122339>
64. Goodman J, Weare J. Ensemble samplers with affine invariance. *Commun Appl Math Comput Sci.* 2010; 5: 65–80. <https://doi.org/10.2140/camcos.2010.5.65>